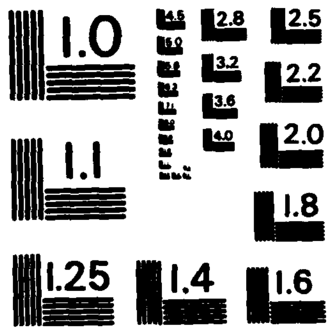
AD-A160 324 HIERARCHICAL MULTISENSOR IMAGE UNDERSTANDING(U) 1/2
HONEYWELL SYSTEMS AND RESEARCH CENTER MINNEAPOLIS MN
R K AGGARWAL ET AL. AUG 85 AFOSR-TR-85-0801
UNCLASSIFIED F49620-83-C-0134 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Hierarchical Multisensor Image Understanding

Final Report

August 1985

Honeywell
Systems and Research Center
2600 Ridgway Parkway
Minneapolis, Minnesota 55413

Air Force Office of Scientific Research
Air Force Systems Command
United States Air Force

SECURITY CLASSIFICATION OF THIS PAGE (WHEN DATA ENTERED)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 85-0801	2. GOV'T ACCESSION NUMBER AD-A160 324	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (AND SUBTITLE) Hierarchical Multisensor Image Understanding		5. TYPE OF REPORT/PERIOD COVERED Final Report October 1983 - August 1985
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(S) Raj K. Aggarwal, Youn Kim, Nelson Marquina, Mike Bazakos, Steve Mader, Firooz Sadjadi, John Budenske, Alianna Maren, Rand Whillock		8. CONTRACT OR GRANT NUMBER(S) F49620-83-C-0134
9. PERFORMING ORGANIZATIONS NAME/ADDRESS Honeywell Systems & Research, 2600 Ridgway Parkway, Minneapolis, Minnesota 55413		10. PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A7
11. CONTROLLING OFFICE NAME/ADDRESS Air Force Office of Scientific Research (AFSC) Bolling AFB DC 20332		12. REPORT DATE August 1985
		13. NUMBER OF PAGES 127
14. MONITORING AGENCY NAME/ADDRESS (IF DIFFERENT FROM CONT. OFF.)		15. SECURITY CLASSIFICATION (OF THIS REPORT) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (OF THIS REPORT)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (OF THE ABSTRACT ENTERED IN BLOCK 20, IF DIFFERENT FROM REPORT)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

Image processing, Image understanding, Artificial intelligence, Scene analysis.

20. ABSTRACT (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

This report describes the research results on Honeywell's Hierarchical Multisensor Image Understanding program. Honeywell is developing a unified framework for the different hierarchical levels of image processing such as segmentation, detection, classification, and identification of outdoor scenes and across different sensor modalities such as millimeter wave, infrared, and visible. Current activities on the project are reviewed under the following headings: (1) A Survey of Multisource Information Fusion Systems; (2) The Role of Structure in Human and Machine Perception; (3) A Knowledge Based Image Segmentation System; (4) The Use of Optical Flow as a Depth Cue in Scene Analysis and (5) Belief Maintenance for a Fuzzy Reasoning System.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (WHEN DATA ENTERED)

Hierarchical Multisensor

Image Understanding

Final Report

Contract F49620-83-C-0134

Honeywell Systems and Research Center

Minneapolis, Minnesota 55413

ABSTRACT

This report describes the research results on Honeywell's Hierarchical Multisensor Image Understanding program. Honeywell is developing a unified framework for the different hierarchical levels of image processing such as segmentation, detection, classification, and identification of outdoor scenes and across different sensor modalities such as millimeter wave, infrared, and visible. Current activities on the project are reviewed under the following headings: (1) A Survey of Multisource Information Fusion Systems; (2) The Role of Structure in Human and Machine Perception; (3) A Knowledge Based Image Segmentation System; (4) The Use of Optical Flow as a Depth Cue in Scene Analysis and (5) Belief Maintenance for a Fuzzy Reasoning System. Past activities on the project which are reported in Annual Report covering period October 1983 - September 1984 include (a) AI-based Generic Image Segmentation and Object Recognition; (b) Evidence-Confidence Paradigms for Image Understanding; (c) Hierarchical Systems Theory for Control Structures, and (d) Invariant Methods in Image Understanding.

A SURVEY OF MULTISOURCE INFORMATION FUSION SYSTEMS

TABLE OF CONTENTS

A SURVEY OF MULTISOURCE INFORMATION FUSION SYSTEMS

1.0 INTRODUCTION

The human brain is able to accept information from multiple sources (i.e. five senses), fuse the information into a single pool of knowledge, reason across the knowledge, predict which of the sources will provide the most beneficial amount of new information, and direct resources to those sources in order to concentrate on processing new information. In this way the human successfully interacts with his environment, assessing and understanding ongoing situations.

Often this information is the result of partial scanning of the domain, resource-limited processing, many-to-less heuristic beliefs, probabilistic assumptions, and dimensional mapping. Thus, human processing uses information which can be incomplete, uncertain, even incorrect, and even inherently evidential. Each piece of information thus has some amount of belief associated with it based on its source, condition of collection, metnod of collection, etc. In other words, the belief is based on evidence. Making an inference about the world based on the beliefs of the information (and associated evidence) requires not only reasoning over the information but reasoning about the belief of the information and the evidence that that belief is based on.

Being able to manage multiple sources of information is a skill at which humans must be proficient in order to interact with their environment. The goal of intelligent systems is also intelligent environmental interaction. These systems will need multiple sources/sensors which can provide a variety of environmental information. Like their human counterparts, intelligent systems which are to interact with their environments must be able to fuse, evidentially reason over, and control the processing of multiple sources of uncertain, complete and incorrect knowledge. What follows is a survey of the design of such systems. In surveying the state-of-the-art in Multisource Information Fusion (MSIF) the research breaks down into three subtopics: The design of the entire MSIF System, the work done in Evidential Reasoning, and in Planning and Control of the multiple sources.

2.0 SYSTEM DESIGN

A Multisource Information Fusion System must perform a multitude of tasks. The sensors must be controlled; the low level data processed into high level knowledge; uncertainty factors assigned to data, knowledge and process/system assumptions; information must be pooled and extrapolated; assumptions about the data must be made and response to these assumptions must be created, planned out and executed. These tasks can be shown by Figure 1. In searching the literature only the work of Garvey and Lowrance [Garvey and Lowrance, 84; Garvey and Fischler,80; and others in the System Design references] has addressed the issue of a complete system design. Their work emphasized a system work listed for battle threat assessment. In their design they use four basic tasks (see Figure 2):

1) ANTICIPATE: Using current information, this task attempts to identify prospective significant events. The system is looking for "What will happen next"? The knowledge used includes known entities in the environment, their capabilities, entities often associated with them, etc. This module is attempting to hypothesize what it is missing or could be about to miss with its sensors. This information is then passed on to the next task.

2) PLAN/ALLOCATE: Given a list of what could be happening in the environment, this module decides what is important to sense and how it will attempt to sense it. A 'plan' is created and passed on.

3) CONTROL: From the plan, this module guides and manages the sensors, parameters, and data operators. The data collected are passed onto the next phase.

4) INTERPRET: The collected data are added to the model of current situation. The new information is inferenced over and the updated world model is passed on to the ANTICIPATE task.

Unfortunately all of Garvey's work past this has been on the INTERPRETATION task where he uses the Dempster-Schafer rules (covered later) to perform evidential reasoning over a hierarchical world model of threat behavior. Garvey assumes that data will be in symbolic form and thus his knowledge representation and fusion mechanisms do not consider any non-symbolic levels of sensory data (ie. image operator results). Possibly these non-symbolic levels of sensory data are unusable until they are transformed into symbolic data. Since Garvey has concentrated on the INTERPRETATION task of his design, the details of the PLANNING/ALLOCATE and CONTROL tasks are unspecified. Much more work needs to be done before Garvey's system can actually be called complete. Though Garvey's work appears to be the only published research concerning a complete system design, many others have been researching the evidential reasoning problem. There appears to be a large void of research in the 'complete' system design and development aspects of Multisource Information Fusion, and though the research on evidential reasoning is necessary, its development must take into consideration the interaction with the remainder of the system. This can only be done after the design of other components has begun.

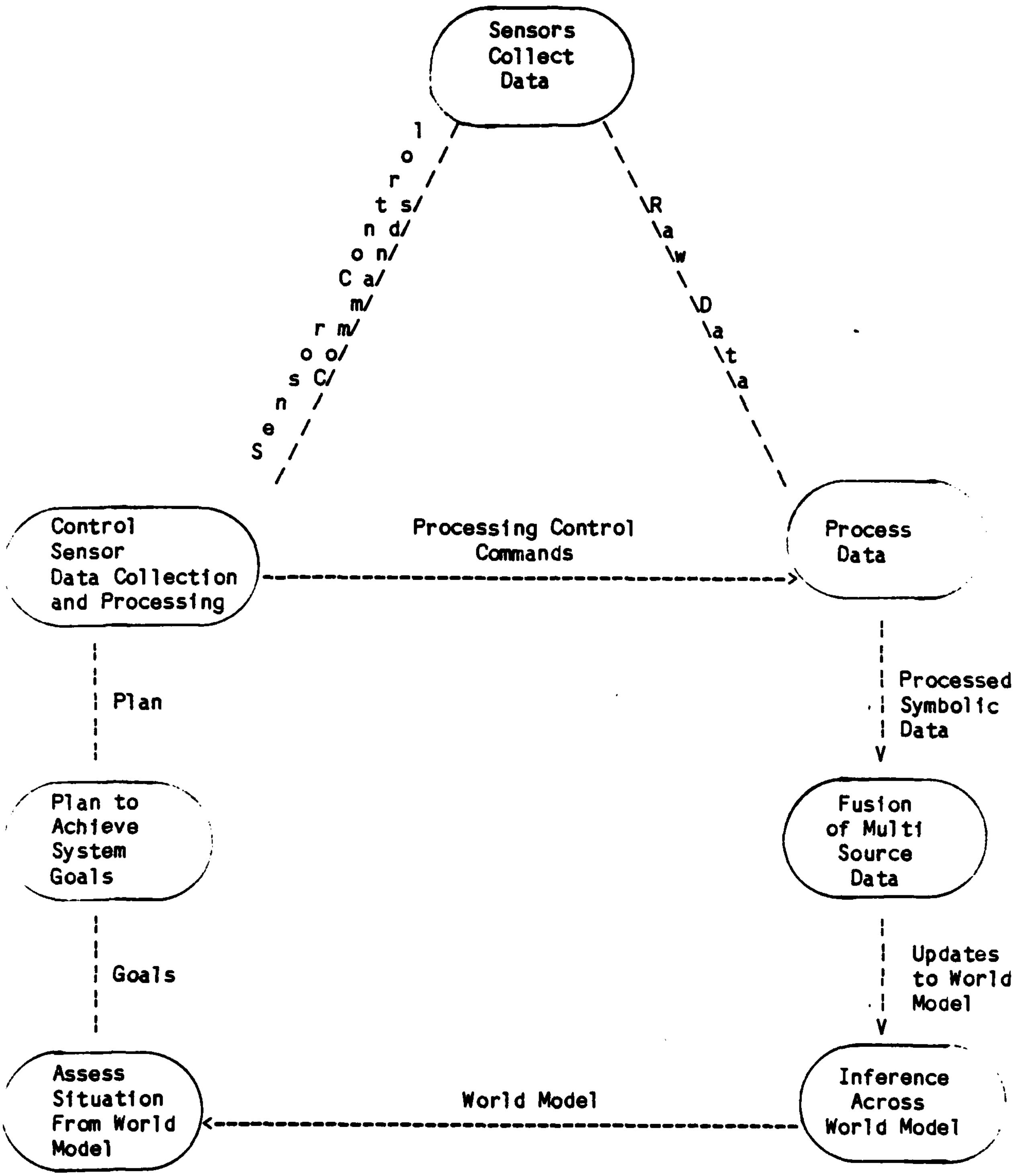


Figure 1: Cycle of Multisensor Information Fusion

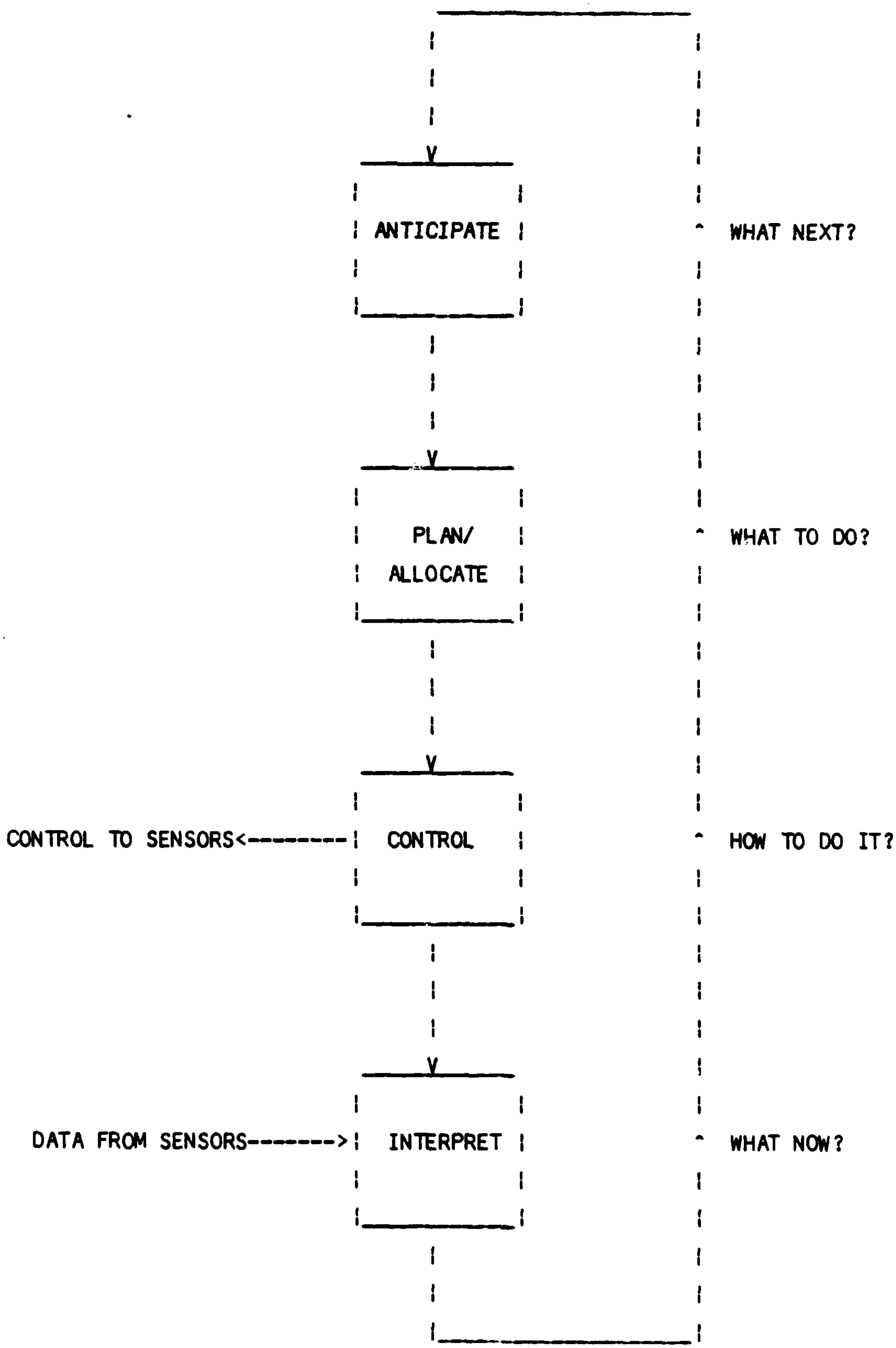


Figure 2: The Basic Loop

3.0 EVIDENTIAL REASONING

3.1 Introduction to Evidential Reasoning

Evidential reasoning has been the main focus of Multisource Information Fusion systems research, because it is a defined subproblem of MSIF systems, and because many feel that it is the heart of the MSIF system. Solving this problem will allow partial implementations of systems and provide a starting point for incremental development. An evidential reasoning system first needs an internal model of the world on which to reason, deduce and inference. Established world model designs (Rich 83) and their inferencing mechanisms are too rigid and inflexible to truly reflect the real world, and though current research efforts are expanding these designs to represent the real world, much of this research has only been at the symbolic level. The issue of how the non-symbolic levels of data are affected by uncertainty, incompleteness, and inconsistency, and whether this effect will affect the remaining mechanism, has yet to be addressed. (This issue is more accurately called the Data Fusion problem but we will use the terms interchangeably).

In order to reflect the real world, data fusion systems must be able to represent the wide range of knowledge found throughout the real world. This knowledge can be declarative or procedural, and due to the conditions under which it was created, it will have an uncertainty associated with it. Uncertainty of the knowledge must be represented, even to the point of being ignorant of the knowledge. The system can be ignorant of a certain fact, and must be able to represent and work around this ignorance of information (however, knowing ignorance of a fact reveals a lack of total ignorance). Finally, when quantities of information are combined into knowledge bases, consistency becomes an issue. It is important also that the facts in the knowledge base do not conflict. Methods such as belief and truth maintenance systems do this kind of consistency checking.

The most important function of a data fusion system is to derive new information from known information. There are two ways to derive new information:

(1) Extrapolation uses procedural rules to derive new information from a single piece of evidence, and

(2) Pooling combines multiple pieces of evidence.

When pooling evidence to form new evidence there can be a problem with interdependencies of different pieces of evidence. For example in a medical diagnosis problem, two symptoms might occur pointing to the same disease. However, this does not always mean a greater chance of the disease than if only one of the symptoms had been observed. It is possible that the two symptoms are dependent and always occur together. Different data fusion systems deal with this dependency problem in different ways. Some even ignore it.

In the following sections three different approaches to data fusion are analyzed and two new theories, each addressing a slightly different problem in evidential reasoning are explained.

3.2 Bayesian Methods

Bayesian probabilities as used in Prospector by Duda et al (Duda, Hart and Nilsson 1976) assigns strict probabilities for each piece of knowledge in the system. The probability of an event H is P(H), where P(H) is a real value between 0 and 1. These probabilities must follow the statistical law that says the sum of all of the probabilities for all possible outcomes of a given event must be 1.

The combination of the associated probabilities when facts are combined is handled by Bayes theorem. Bayes theorem says that the probability of an event H based on some observed evidence E is given by:

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Likewise,

$$P(H|E) = \frac{P(E|H) * P(H)}{\sum_{i=1}^{K} P(E|H_i) * P(H_i)}$$

where:

$P(E|H_i)$ = the probability of the evidence E given H_i.

$P(H)$ = the apriori probability of the event H.

K = the number of different possible outcomes for H.

There are some problems with using Bayes theorem for real world applications however. It requires knowledge of all of the conditional probabilities, and this information is often not available. Bayes theorem has limited application in a world where we do not have access to the probabilities for all of the events we are concerned with.

The Bayesian method also has problems with dependence of certain facts. By definition Bayes theorem requires all evidences to be independent. This is often not true in the real world.

Despite their shortfalls, Bayesian methods can be useful in limited problem domains where much information about the probabilities of events are known. Mineral exploration, which is the domain for Prospector, is one such area.

Charniac (Charniac, 1983) points out that for many expert systems applications such as medical diagnosis, Bayesian methods can be used without the independence assumption. He shows that the interdependence of symptoms will affect the probabilities for all diseases equally and will therefore not change the relative rankings of the diseases. He goes on to state that other dependency problems can be addressed by combining the dependent evidences into single states.

There are many systems that use derivatives of Bayesian for data fusion. References to many of these are listed in the bibliography.

3.3 MYCIN

In MYCIN, Shortliffe (Shortliffe, 1975) attempted to overcome the shortcomings of Bayesian approaches while retaining the advantages, by using a system that was an approximation to conditional probabilities. Each piece of information in MYCIN has associated with it a value between 0 and 1 which is its belief value: $P(H)$. The disbelief value is $1-P(H)$. Any assertion about this fact has two measures associated with it. A measure of belief (MB) is the measure of the decrease in disbelief of H as a result of a piece of evidence and a measure of disbelief (MD) is a measure of the decrease in belief of H as a result of the evidence. MB is defined by:

$$MB = \begin{cases} 1 & : \text{if } P(H) = 1 \\ \dfrac{MAX[P(H|E), P(H)] - P(H)}{1-P(H)} & : \text{otherwise} \end{cases}$$

and MD is defined by:

$$MD = \begin{cases} 1 & : \text{if } P(H) = 0 \\ \dfrac{MIN[P(H|E), P(H)] - P(H)}{-P(H)} & : \text{otherwise} \end{cases}$$

In addition, a certainty factor is computed which is the measure of belief minus the measure of disbelief.

The above formulas handle extrapolation where a single piece of evidence leads to a conclusion. To pool multiple pieces of evidence MYCIN uses the following formulas:

$$MB[H|S_1+S_2] = \begin{cases} 0 & : \text{if } MD[H|S_1+S_2] = 1 \\ MB[H|S_1] + MB[H|S_2] * (1-MD[H|S_1]) & : \text{otherwise} \end{cases}$$

and,

$$MD[H|S_1+S_2] = \begin{cases} 0 & : \text{if } MB[H|S_1+S_2] = 1 \\ MD[H|S_1] + MD[H|S_2] * (1-MD[H|S_1]) & : \text{otherwise} \end{cases}$$

The reasoning system used in MYCIN allows uncertainty and allows us to represent the lack of evidence for a certain conclusion. For example, if the evidence does nothing to confirm a hypothesis then MB=0. If a piece of evidence does nothing to disprove a conclusion then MD=0. This method of reasoning does, however, still have some shortcomings. We still have dependency problems just as we did with the Bayesian methods but now we also have an added disadvantage in that the measures of belief and disbelief are not true probabilities and therefore cannot be expected to follow the laws of probabilities. These values are arbitrary, and at some level are assigned by a human. This poses a problem when information from different humans is used in the same system. Invariably they will use different scales in assigning values.

3.4 Dempster-Schafer

The Dempster-Schafer theory (Barnett, 1981) uses the same interval concepts as MYCIN but extends them so that the probabilities are represented by mass distribution functions. This theory provides rules for assigning and manipulating these distributions. The intervals are defined as follows:

[S(A), P(A)]

where: S(A) is the degree of support,

P(A) is the degree of plausibility or the degree of failure to refute,

P(A)-S(A) is the degree of ignorance.

These values are computed using a mass distribution function which distributes a belief value over the entire range of possible hypotheses. The value for S(H) or the degree of support for a given hypothesis H is equal to the mass distribution function summed over all the hypotheses that imply H:

$$S(H) = \sum_{E_i \subseteq H} M(E_i)$$

The value for P(H) or the plausibility of H is the sum of all of the hypotheses that do not imply not (H):

$$P(H) = \sum_{E_i \cap H <> 0} M(E_i) = 1 - S(\bar{H})$$

To pool multiple pieces of evidence to support one hypothesis we use Dempster's rule for combination:

$$M(H) = \frac{1}{1-K} \sum_{E_i \cap E_j = H} M_1(E_i) * M_2(E_j)$$

where

$$K = \sum_{E_i \cap E_j = 0} M_1(E_i) * M_2(E_j)$$

The Dempster-Schafer thecrem, because it follows the law of probability, gives us many nice properties. The combination rule is commutative and associative so evidence can be combined in any order or grouping. Also, when the probabilities are known exactly (when S(H) = P(H)) the Dempster-Schafer law reduces to Bayes theorem.

The Dempster-Schafer theorem fills many of the needs for an information combination system and is widely used today. References to many systems that use the Dempster-Schafer theorem are listed in the bibliography.

3.5 Edge Merits

Slagle (1984) in his battle system proposes an extension to the Bayesian and MYCIN combination systems that takes into consideration the probability associated with the rule as well as the evidence. These probabilities are called edge merits, and are used to propagate the values of evidence through rules as well as heuristic in rule conflict resolution. The edge merit for an AND combination is defined as:

$$\frac{P(H)}{P(E)}$$

The edge merit for an OR combination is defined as:

$$\frac{1-P(H)}{1-P(E)}$$

Then, when the rule is fired, the AND function returns the minimum of the probabilities of its arguments and the OR function returns the maximum of its arguments.

Edge Merits can be incorporated into a data fusion scheme that uses any of the above methods.

3.6 Endorsements and Endorsers

Cohen and Grinberg (Cohen & Grinberg, 1983) developed a theory of heuristic reasoning about uncertainty which is symbolic (non-numerical) and uses some of the concepts from Doyle's Truth Maintenance System. They propose a system in which evidence for an inference is associated with that inference and is called an Endorsement for that inference (much like Justifications in a Truth Maintenance System). Thus, endorsements are records of the inferences which have taken place, and Endorsers are defined as the computations that assert these records. (See section on Non-monotonic Reasoning.) Unfortunately, no further work on implementing a system based on the theory has been published. Nonetheless, the theory is interesting enough to be briefly explained below.

Cohen and Grinberg claim that the numerical approaches to reasoning under uncertainty restrict the amount of heuristic knowledge about uncertainty and evidences, knowledge that humans actually use. In numerical methods, a number is merely a summary of the evidence which supports an inference, and actual evidence is left unaccessible by the processes reasoning about future inferences. There are two reasons why the evidence for an inference should be accessible. First, a number cannot relate the type of evidence which supports the belief. Some types of evidences will have more importances in certain situations or contexts (i.e. corroborative evidence vs. contradictory evidence), and knowing what kind of evidence is supporting an inference can only aid the reasoning process. The second reason to make the evidence accessible is the belief of an inference is based on the context of its evidence, which includes the current inference being reasoned about. Therefore, evidence used for one inference is in a different context from evidence used for another inference, and thus the belief in the evidence itself can be different between the two contexts. This becomes most important when dealing with evidence which supports an inference being used as evidence for a new inference (i.e. propagation; numerical methods can often yield meaningless, and irrelevant belief values through propagation).

The certainty of an inference is represented by the strongest endorsement for the inference. Therefore, an inference supported by one kind of endorsement (eyewitness evidence) would have a higher certainty than if the inference was supported by less preferred evidence (circumstantial evidence). This means that knowledge is needed to define and rank a characteristic (or primitive) set of domain endorsements. Also, knowledge is needed to heuristically propage endorsements over inferences (much like degrees of belief are numerically propagated over inferences (much like degrees of belief are numerically propagated over inferences by combining functions), but the propagaton must be sensitive to the context of the inference. Rules are needed to propagate endorsements over inferences, thus serving the same purpose as combining functions, with each domain of expertise having numerous idiosyneratic rules for covering special cases of endorsement propagation.

The Endorsement Theory is a fresh look at reasoning with uncertainty, but the theory leaves many unanswered questions. The knowledge and structure of the rules is unclear, as well as the final form of each inference. These holes will have to be addressed before the theory can be implemented into a working system. One benefit of the theory is that by retaining the endorsements of an inference, one can discount the uncertainty of the evidence once the use of the inference is known (i.e. discounting the uncertainty of one value when it is averaged with other values). In this manner, a better grip on the propagation of uncertainty is maintained, thus yielding a more understandable belief of the inferences.

3.7 <u>Non-Monotonic Logic</u>

Traditional reasoning systems based upon predicate logics are considered monotonic in that the number of statements known to be true is strictly increasing over time. All newly added statements and newly proven theorems cannot disprove any of the previously known knowledge in the system. Unfortunately the world is not monotonic; but is incomplete, constantly changing, and in order to reason efficiently over complex problems, assumptions (default reasoning) must be made about the world which can possibly be proven incorrectly later in the reasoning process. From this, the development of Non-Monotonic Logics has begun [Doyle and McDermott, 1979, McDermott and Doyle, 1980], and though the theory is fairly new, its basic concepts allow default-reasoning assumptions to be made and retracted without disrupting the belief integrity of the world model.

Default reasoning allows the inclusion of logical statements on the order of "If X cannot be proven with you have right now, then conclude Y" in the problem solving process. Thus assumptions can be made and considered true until proven wrong. When a contradiction is found, backtracking is performed to the assumption which caused the contradiction and then that statement and all statements derived from it are withdrawn from the world model. This is called dependency-directed backtracking [Stallman and Sussman, 1977].

Non-monotonic reasoning systems have two added components over common inferencing systems. The first is an Assumption mechanism which creates assumptions based usually on defaults about partial solutions to aid in solving the problem. The second component is a truth maintenance mechanism which upon the discovery of a contradiction finds the faulting statement and revises the truth of the world model by withdrawing the statement and all of its dependents from the world. Most of the work in building non-monotonic reasoning systems has been based on Doyle's Truth Maintenance Systems (TMS) [Doyle 1979, a&b] which is an implemented system that supports non-monotonic reasoning by serving as a truth maintenance subsystem available to other reasoning programs. The system does not generate new inferences but maintains the integrity and consistency across the statements produced by the reasoning program via its own dependency-directed backtracking mechanism. Since TMS is the basis of most research in non-monotonic reasoning, it is explained below.

In TMS, each statement is called a 'node' and is, during the reasoning process, either believed to be true (IN) or not believed to be true (OUT). OUT statements are not believed because there exists no reason for believing them, or because none of the possible reasons for believing are currently true. Associated with each node is a list of justifications which reflects how the validity of one node can depend on the validity of others, and of which these are two kinds: Support Lists [SL (in-nodes) (out-nodes)] and Conditional Proof [CP (consequent) (in-node) (out-node)]. Supports Lists are the most common and its node is IN if all of the in-nodes are IN and all of the out-nodes are OUT. If there are no in-nodes nor out-nodes then the statement is considered a 'premise.' Conditional Proofs are hypothetical arguments which hold derived contradictions within the world.

As an example, suppose we are inferencing as to what type of animal IVAN is. We will start out with:

N1.	Animal(IVAN) = Shark	IN	[SL ( ) (N2)]
N2.	Animal(IVAN) <> Shark	OUT	
N3.	Animal(IVAN) <> Dog	OUT	

The system assumes that IVAN is a shark (node 1 is IN) since there is no reason to believe it isn't. From these the system could infer that since Ivan is a shark, then he can also swim:

N4.	IVAN can swim	IN	[SL (N1) ( )]

To show the maintenence system in action, assume that new information is introduced by statements N25 and N32 (from else where in the system) and found to contradict:

N5.	IVAN has fur	IN	[SL (N25, N32) ( )]

The inference system realizes a contradiction in that sharks do not have fur and so the dependency directed backtracking is triggered. In searching back, the system realizes that N1 is an assumption and could be causing the contradiction. First, the contradiction is marked:

N6.	CONTRADICTION	IN	[SL (N1, N5) ( )]
N7.	NO-GOOD N1	IN	[CP N6 (N1, N5) ( )]

The NO-GOOD node marks the proof that if N1 IVAN is a shark and N5 IVAN has fur then N6 we have a contradiction. Without N7, we would never know that we tried N1 if we ever have to backtrack again. Next, we update the rest of the world model and select a new assumption:

N1.	Animal(IVAN) = Shark	OUT,	[SL ( ) (N2, N3)]
N2.	Animal(IVAN) <> Shark	IN	[SL (N6) ( )]
N3.	Animal(IVAN) = Dog	IN	[SL (N2) (N8)]
N4.	IVAN can swim	OUT	[SL (N1) ( )]
N5.	Animal(IVAN) <> Dog	OUT	
N6.	CONTRADICTION	OUT	[SL (N1, N5) ( )]

Now the inference mechanism can proceed and derive new deductions, assumptions and contradictions.

There are a number of variations to the TMS scheme [London 1978; Thompson, 1979; Ginsberg, 1984] but all are based on the same theoretical concepts and all offer considerable improvements over other classical systems. Since the inferencing and dependency directed backtracking is non-chronological, the support relationships rather than the temporal orderings determine the recovery from an error and thus the appropriate erronious assumptions are found quickly. Also the use of Conditional Proof (CP) structures allows the causes of contradictions to be summarized and recorded, thus mistakes are made only once. These two improvements provide an increase in efficiency large enough to offset the overhead of maintaining the justifications (though overhead would depend on implementation.) Difficulty in using TMS could develop when marking an assumption, (i.e. IVAN is a shark) when all other possibilities must be created and accounted for in the out-list of the assumption:

N1.	Animal(IVAN) = Shark	IN	[S1 ( ) (N2, N3...n)]
N2.	Animal(IVAN) <> Shark	OUT	
N3.	Animal(IVAN) <> dog	OUT	
N4.	Animal(IVAN) <> Platypus	OUT	

Possibly mechanisms can be created to introduce alternative assumptions, one by one, as they are needed. This would greatly reduce overhead.
Using the Non-monotonic Logic provides great advantages over classical logics and is an excellent augmentation to uncertainty inferencing systems. To date, no one has yet published an implementation of a truth maintenance system complete with uncertainty factors for its assumptions. Possibly, the establishment of both (non-monotonic and uncertainty) theories will allow development of an integrated theory. For a brief view of TMS see [Rich, 1983, Doyle, 1979a], and see [Doyle, 1979b] for an extended view. For the mathematical background on Non-monotonic Logic, see [McDermott and Doyle, 1980; Doyle, and McDermott 1979c,].

3.8 Fuzzy Sets and Logic

Fuzzy set theory is a well-developed mathematical theory which has not yet been significantly exploited in A.I. systems. It provides an augmentation to evidential reasoning systems which must reason with fuzzy qualified statements such as: "most extremely fat women are very sweaty", where "most", "fat", "sweaty", "extremely", and "very" use fuzzy qualities (not exact boundry between membership/non-membership of the quality.)

The theory of fuzzy sets is the development of a body of concepts and techniques for systematically dealing with imprecise boundaries between classes of objects. A fuzzy set is a class in which there is gradual progression from membership to non-membership and each object in the set has a grade of membership intermediate between 1 (full membership) and 0 (non-membership). Thus a conventional set is a degenerate case of a fuzzy set where only two grades of membership are allowed: 1, 0.

The need of a fuzzy logic is due to the fuzzyness of the world. We often use qualifiers in describing membership to a set such as: very, quite, almost, slightly. In order for a system to reason about fuzzyness it must first represent it. Let U be our universe of discourse (e.g. set of integers) and A be a fuzzy subset of U, which is characterized by a membership function:

$$v_A = U \rightarrow [0, 1]$$

which associates with each element u of U a number $v_A(u)$ in the interval [0, 1], with $v_A(u)$ representing the grade of membership of u in A.

<u>Example</u>: Let the universe of discourse be the set: U = [1, 2, 3, 4, 5...] with u interpreted as "small". A fuzzy subset of U labeled "very" may be defined as:

$$\text{very small} = (v_1:u_1) + (v_2:u_2) + (v_3:u_3) + (v_4:u_4) + (v_5:u_5)$$
$$= 1:1 + 1:2 + 0.5:3 + 0.25:4 + 0.05:5$$

where : is a separator to avoid confusion and + means union of the elements.

The members: 1, 2 of "small" have a grade of 1, while the member 3 has a grade of 0.5 (i.e. not as very small as 1 or 2 are, and so on.) From here a number of theoretical concepts can be defined including classical set definitions (i.e. containment) set operations (i.e. complement, union), relationships and other principles. For a more thorough account of fuzzy set theory see [Zadeh, 1977].

Fuzzy logic is an extension of fuzzy set theory which provides a representation for fuzzy quantifiers and truth-values, as well as provides a set of translation and inference rules which can reason over the fuzzy representation. Fuzzyness is represented through possibility distributions T_x = the possibility distribution of X) which is a heuristically derived measure of the semantic fuzzyness of linguistic variables. Let F be a fuzzy subset of U = {0, 1, 2...} and "X is F" represents F = T_x, or:

"X is F -> T_x = F"

An example:

Let

X is small -> T_X = 1:0 + 1:1 + 0.8:2 + 0.6:3 + 0.4:4 + 0.2:5.

Then

Poss{X = 0} = 1

Poss{X = 1} = 1

⋮

Poss{X = 5} = 0.2

Where Poss{X = u} = $v_F(u)$ is the possibility that X may take u as a value. So if I say "there is a small number of people left", the possibility that there is only 2 left is Poss{X = 2} = $v_F(2)$ = 0.8. Possibility is different from probability in that probability is a measure of randomness while possibility is a measure of semantic fuzzyness or imprecision in value. A variable in fuzzy logic is considered a linguistic variable whose values are represented as words or sentences in a natural or synthetic language. The value of each variable defines a possibility distribution in the domain of the variable. For example: given the primary variable TRUE, its antonym FALSE, and a finite set of modifiers and connectives such as <u>and</u>, <u>or</u>, <u>not</u>, <u>very</u>, <u>more or less</u>, <u>extremely</u>, etc... the linguistic value of TRUE may be generated and represented as:

```
True                          False
not true                      not false
very true                     very false
not very true                 not very false
more or less true             more or less false
   ⋮
not true and not false
not very true and not very false
   ⋮
```

The linguistic truth-value is a composition of possibility distributions of the primary variable and attached modifiers. Translation rules provide a means of deriving the composite truth-values. Translation rules fall into the four categories as explained below.

1. Modification rule (not, very, more or less, etc.)

If X is F -> $T_x = F$

Then X is mF -> $T_{mx} = F^+$

where m is a modifier of F and F^+ is a modification on F.

<u>Example</u>: Let F = small, T_x =1:0 + 1:1 + 0.8:2 + 0.6:3 + 0.4:4 + 0.2:5.

Let m = very, and $F^+ = (F)^2$ (F squared)

Then:

X is very small --> $T_{mx} = (F)^2$ =

1:0 + 1:1 + 0.64:2 + 0.36:3 + 0.16:4 + 0.04:5

2. Conjunctive, disjunctive and implicational rules: Let F and G be fuzzy subsets of U and W,

X is F --> $T_x = F$ and Y is G --> $T_y = G$

a. X is F and Y is G --> $T_{(x,y)} = F \times G$ where:

$$v_{(F \times G)}(u,w) = \min(v_F(u), v_G(w))$$

b. X is F or Y is G --> $T_{(x,y)} = \overline{F} \cup \overline{G}$ (union)

where:

$\overline{F} = F \times W$ (compliments)

$\overline{G} = G \times U$

$v_{(\overline{F} \cup \overline{G})}(u,w) = \max(v_F(u), v_G(w))$

c. If X is F then Y is G --> $T_{(x/y)} = \overline{F} \oplus \overline{G}$

where $T_{(x/y)}$ is the conditional possibility distribution of Y given X and

$$v_{(\overline{F} \oplus \overline{G})}(u,v) = \mathrm{Min}(1, (1 - v_F(u) + v_G(w)))$$

<u>Example</u>: Let

$$F = \text{SMALL} = 1{:}1 + 0.6{:}2 + 0.1{:}3$$
$$G = \text{LARGE} = 0.1{:}1 + 0.6{:}2 + 1{:}3$$

Then

X is small and Y is large = $T_{(x,y)}$

$$= 0.1{:}(1,1) + 0.6{:}(1,2) + 1{:}(1,3) + 0.1{:}(2,1) + 0.6{:}(2,2) + 0.6{:}(2,3) + 0.1{:}(3,1) + 0.1{:}(3,2) + 0.1{:}(3,3)$$

3. Quantification rule (many, few, several, all, some, etc...): If $U = \{u, \ldots u_n\}$, Q is a quantifier and

$$F = v_1{:}u_2 + v_2{:}u_2 + \ldots + v_n{:}u_n$$
$$X \text{ is } F \dashrightarrow T_x = F$$

then "QX are F" (eg. "several X's are large") translates to

$$T_{count(F)} = Q$$

where

$$T_{count(F)} = \sum_{i=1}^{N} v_i$$

<u>Example</u>:

Let

$$\text{SEVERAL} = 0{:}1 + 0.4{:}2 + 0.6{:}3 + 1{:}4 + 1{:}5 + 1{:}6 + 0.6{:}7 + 0.2{:}8$$

Then SEVERAL X's are LARGE = $T_{\sum_{i=1}^{N} v_{LARGE}(u_i)}$

$$= 0{:}1 + 0.4{:}2 + 0.6{:}3 + 1{:}4 + 1{:}5 + 1{:}6 + 0.6{:}7 + 0.2{:}8,$$

where $v_{LARGE}(u_i)$ is the grade of membership of the i^{th} value of X in the fuzzy set LARGE.

4. Truth qualification rule: Let t be a fuzzy truth-value like "very true", "quite true", etc. Then "It is t that X is F" is expressed as:

$$X \text{ is } F \text{ is } t \dashrightarrow T_X = F^+$$

where

$$v_{F^+}(u) = v_t(v_F(u))$$

<u>Example</u>:

Bob is young is very true

where:

$$m = \text{very} \dashrightarrow F^2$$

then

$$T_{age(Bob)} = v_{true^2}(v_{young}(u)),$$

where u is an element in the interval [0,100]

assuming

$$v_{young}(u) = \left(1 + \left(\frac{u}{25}\right)^2\right)^{-1}$$

and

$$v_{true}(w) = W^2, \quad w \text{ is an element in the interval } [0, 1]$$

then

$$T_{age(Bob)} = \left(1 + \left(\frac{u}{25}\right)^2\right)^{-4}$$

The translation rules above can be combined to provide the possibility distribution of composite propositions. From here, rules for inferencing over the possibility distributions can be defined. Zadeh [Zadeh 79] identifies complex rules for projection and conjunction and combines them into a composite rule of inference which is a generalized version of the classical modus ponens. Thus he claimed to be able to infer the possibility distribution of Y from the knowledge of X's possibility distributions and from the conditional possibility distribution of Y given X. It is through this composite rule of inference that Zadeh claims will expand the applicability of rule-based systems, giving them an interpolative (Fuzzy) capability. Fuzzy logic is not an alternative to the other work in this survey, but is an augmentive representation which can be integrated with non-monotonic logic and evidential (uncertainty) reasoning. All of these theories are complex and implementing an integrated system of them may be close to impossible. Only after they are individually developed will that question be answered. For the best coverage of fuzzy set and logic theory see [Zadeh, 1977, Zadeh, 1979].

4.0 PLANNING AND CONTROL

A plan is a partially ordered net of operations, each performable by a host unit. Planning is the process of ordering the operators such that their actions and resource needs do not conflict. In a Multisource Information Fusion System there can be potentially high data rates, and thus sensor resources and data processing controls needs to be managed effectively, with the collection of important information being optimized and irrelevant data filtered out. Decisions must be made as to what sources (of information) on which to concentrate; what processes to run; when and where to run them; what parameters to use; and which data to pass along. Memory, bus, processors and time must all be allocated. The entire information collection process must be planned out such to optimize the collection of vital information. Though Garvey and Lowrance [Garvey and Lowrance, 1984] promote the instigation of planning within a MSIF system, current planning strategies have yet to be adopted for MSIF designs. Garvey has looked at the use of planning in vision systems [Garvey, 1976, Ballard and Brown, 1982] but his approach is very high level (i.e. searching for a telephone in an office by first planning to find the table and searching the table top). Very little work has been done in applying automated planning strategies to plan the control of large MSIF systems.

The current state-of-the-art in Automated Planning uses a hierarchy of goals to create a plan [Sacerdoti, 1977], and meta-level knowledge to control the plan creation [Hayes-Roth et al, 1979, Stefik, 1981]. Given a high level goal with constraints, the goal is expanded into a partially ordered net of (children) subgoals and actions which will achieve the (parent) goal. Then the subgoals are examined and compared to an internal world model to see if it is already true. If not, the net is examined to see if restructuring the plan will make the subgoal true and if not, the subgoal is made true by expanding it into more subgoals. This process continues until all goals are true. Intermitently the plan-net is examined for conflicts between goals (such as attempting to use the same resources at the same time). Often conflicts are resolved by reordering the plan-net. Goals are kept in parallel until a necessary ordering can be determined [Sacerdoti, 1977], and variables are not arbitrarily bound but have constraints (which describe the value) placed on them until a correct value is found [Stefik, 1981 a&b].

The processes involved in planning are quite simple; the difficulty is in placing knowledge about the host unit's (in MSIF a source system controller) actions into a planning system operator. The effects prerequisites constraints on variable bindings of when, how, and why (intentions) to use the action and much more must be acquired from all possible actions of the host unit. Incomplete or incorrect knowledge can cripple the planning process.

At the end of the planning process the final plan can progress through a number of optimization and interpretation processes [Kempf, 1983] which will transform the plan into a program for the controller to follow. Possibly the controller will have these processes built into it and will intelligently use the plan as a knowledge base which guides the controller's actions [Budenske, 1984].

The duty of the controller is to manage the operation of the sensors. With the high flux of possible data to be collected, it is desirable to focus the sensors on portions of the external world where pertinent information is more likely to be extracted. Not only can the high data rate be cut by selective sensing but also by selective processing of the raw data. The controller will receive a plan from the planning module and will follow this plan allocating system resources managing sensors executing processes and passing the resulting knowledge on to the data fusion module. Possibly, the controller will also control the data fusion and inferencing processes through assignment of initial probabilities:uncertainties and other parameter adjustments. The exact methods of planning and control have yet to be fully examined and thus provide an excellent area for fruitful extension of the state-of-the-art.

5.0 SUMMARY

The research in Multisource Information Fusion systems is very lacking. The majority of work has been in the area of evidential reasoning with the tasks of planning and control being ignored as well as the overall design of an integrated system. The problems in evidential reasoning have been approached through different inferencing mechanisms and logical theories, all of which are still in the early periods of development. The integration of some of the mechanisms and logical theories could greatly increase the evidential reasoning capabilities of a MSIF system, but these mechanisms and theories must be furthered developed before integration can occur.

References

Ballard, D.H., and Brown, L.M., Computer Vision, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1982, pp. 453-461.

Barnett, J., Computational Methods for a Mathematical Theory of Evidence, International Joint Conference on Artificial Intelligence Proceedings, 1981.

Budenske, J., Selected Issues in the Design of a Planning System, Honeywell SRC Internal Report, Dec. 1984.

Charniak, E., The Bayesian Basis of Common Sense Medical Diagnosis, Proceedings of National Conference on Artificial Intelligence (AAAI - 83), 1983.

Cohen, P.R., Grinbery, M.R., "A Theory of Heuristic Reasoning About Uncertainty",, AI Magazine, Vol. 4, No. 2, 1983, 17-24

Doyle, J., A Glimpse of Truth Maintenance, International Joint Conference on Artificial Intelligence Proceedings, 1979, pp. 232-237.

Doyle, J., A Truth Maintenance System, Artificial Intelligence, Vol. 12, #3, 1979.

Doyle, J., and McDermott, D., An Introduction to Non-Monotonic Logic, International Joint Conference on Artificial Intelligence Proceedings, 1979,. pp. 562-567.

Duda, R., Hart, P., and Nilsson, N., Subjective Bayesian Methods for Rule Based Inference Systems, National Computer Conference Proceedings, 1976.

Fischler, M., Tanenbaum, J., and Wolf, H., Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using A Multisource Knowledge Integration Technique, SRI Technical Note #200, Dec. 1979.

Garvey T., and Fischler, M., The Integration of Multisensor Data for Threat Assessment. ICPR Proceedings, 1980, pp. 343-347.

Garvey, T., Lowrance, J., and Fischler, M., An Inference Technique for Integrating Knowledge From Disparate Sources, International Joint Conference on Artificial Intelligence Proceedings, 1981, pp. 319-325.

Garvey, T., and Lowrance, J., An AI Approach to the Integration of Information, SRI Technical Note #318, April 1984.

Garvey, T.D., "An Experiment with a system for Locating Objects in Multisensory Images", Third International Joint Conference on Pattern Recognition, 1976, pp. 567-575.

Ginsberg, M., Non-Monotonic Reasoning Using Dempster's Rule, Proceedings of National Conference on Artificial Intelligence, 1984.

Hayes-Roth, B., Hayes-Roth, F., Rosenschein, S. and Cammarata, S., Modeling Planning as an Incremental Opportunistic Process, International Joint Conference on Artificial Intelligence, #6, Vol. 1, 1979.

Ishizuka, M., Fu, K., and Yao, J., Inexact Inference for Rule Based Damage Assessment of Existing Structures, International Joint Conference on Artificial Intelligence Proceedings, 1981.

Kempf, Karl G., "Artificial Intelligence Applications in Robotics - A Tutorial", a special tutorial at International Joint Conference on Artificial Intelligence, #8, 1983.

Konolige, K., An Information-Theoretic Approach to Subjective Bayesian Inference in Rule-Based Systems, Draft, 1982.

Lowrance, J., and Garvey, T., Evidential Reasoning: An Implementation for Multisensor Integration, SRI Technical Note #307, Dec. 1983.

Lu, S., and Stephanou, H., A Set-theoretic Framework for the Processing of Uncertain Knowledge, Proceedings of National Conference on Artificial Intelligence, 1984.

McDermott, J., R1: A Rule Based Configurer of Computer Systems, AI Vol. 19, Sept. 1982.

McDermott, D., and Doyle, J., "Non-Monotonic Logic I", Artificial Intelligence, Vol. 13, Apr. 1980.

Mizumoto, M., Fukami, S., and Tanaka, K., Fuzzy Conditional Inferences and Fuzzy Inferences with Fuzzy Quantifiers, Internationl Joint Conference on Artificial Intelligence Proceedings, 1979, pp. 589-591.

Pednault, E., Zucker, S., and Muresan, L., On the Independence Assumption Underlying Subjective Bayesian Updating, Artificial Intelligence 16, 1981.

Rich, E., Artificial Intelligence, McGraw-Hill, Inc., 1983.

Rosenberg, S., Reasoning in Incomplete Domains, International Joint Conference on Artificial Intelligence Proceedings, 1979, pp. 735-737.

Sacerdoti, E.D., Planning in a Hierarchy of Abstraction Spaces, Artificial Intelligence, Vol. 5, 1974.

Sacerdoti, E.D., A Structure for Plans and Behavior, Elsevier North-Holland, New York, 1977.

Shortliffe, E., and Buchanan, B., A Model of Inexact Reasoning in Medicine, Mathematical Biosciences 23, 1975.

Slagle, J., Gaynor, M., and Halpern, E., An Intelligent Control Strategy for Computer Consultation, IEEE PAMI 6, No. 2, March, 1984.

Stallman, R., and Sussman, G., Forward Reasoning and Dependency-directed Backtracking in a System for Computer Aided Circuit Analysis, Artificial Intelligence, Vol. 9, #2, 1977.

Stefik, M., Planning and Meta-Planning (Molgen: Part 2), Artificial Intelligence, 16, 1981, pp. 141-170.

Stefik, M., Planning with Constraints (Molgen: Part 1), Artificial Intelligence, Vol. 16, 1981, pp. 111-140.

Strat, T., Continuous Belief Functions for Evidential Reasoning, Proceedings of National Conference on Artificial Intelligence, 1984.

Thompson, A., Network Truth Maintenance for Deduction and Modeling, International Joint Conference on Artificial Intelligence Proceedings, 1979, pp. 877-879.

Umano, M., Mizumoto, M., and Tanaka, K., A System for Fuzzy Reasoning, International Joint Conference on Artificial Intelligence Proceedings, 1979, pp. 917-919.

Zadeh, L., Theory of Fuzzy Sets, Encyclopedia of Computer Science and Technology, Marcel Dekker, New York, 1977.

Zadeh, L., Approximate Reasoning Based on Fuzzy Logic, International Joint Conference on Artificial Intelligence Proceedings, 1979, pp. 1004-1010.

Zadeh, L., Making Computers Think Like People, IEEE Spectrum, Aug., 1984.

TR85.1

THE ROLE OF STRUCTURE IN HUMAN AND MACHINE PERCEPTION

ABSTRACT

For some time, there has been a growing awareness among the Image Understanding (IU) research community that the traditional approaches were not yielding satisfactory results in terms of desired performance. This is strongly related to the emphasis in traditional IU on segmenting and characterizing distinct regions or edge segments. Relationships between regions have not received much attention.

In contrast, researchers in human perceptual processes have long been aware of the importance of the use of relationships between regions, leading to grouping of regions. This capability to form groups of highly related regions is a fundamental (low-level) form of structuring the information in an image. Further work has shown that humans make extensive use of the symmetry properties of image configurations in building up internal symbolic representations of the perceived images. It has become apparent that this capability is not a trivial one, and that human facility in working with structural groupings and symmetric relations develops only in the latter stages of childhood.

These factors provide strong argument that one of the major needs of image understanding systems now is a robust, generic method for representing and processing both group-oriented and symmetry-oriented structural properties of images. This paper illustrates a method for representing the low-level (grouping) structure of segmented images. The structuring process is thoroughly based on an implementation of the factors which a dominant role in human perceptual grouping processes: similarity, proximity, containment, and similar directionality.

The implementation scheme has been applied to both natural (FLIR) and artificial (Bongard) images. The resulting Hierarchal Relational Structures (HRS) provide an organizing schema for grouping related regions for further processing. The HRS further enables a potentially simpler form of representation for higher-order relationships (such as symmetry), and provides a succinct structure to which meaning or object/feature identification processes can be applied.

TABLE OF CONTENTS

THE ROLE OF STRUCTURE IN HUMAN AND MACHINE PERCEPTION

ABSTRACT

For some time, there has been a growing awareness among the Image Understanding (IU) research community that the traditional approaches were not yielding satisfactory results in terms of desired performance. This is strongly related to the emphasis in traditional IU on segmenting and characterizing distinct regions or edge segments. Relationships between regions have not received much attention.

In contrast, researchers in human perceptual processes have long been aware of the importance of the use of relationships between regions, leading to grouping of regions. This capability to form groups of highly related regions is a fundamental (low-level) form of structuring the information in an image. Further work has shown that humans make extensive use of the symmetry properties of image configurations in building up internal symbolic representations of the perceived images. It has become apparent that this capability is not a trivial one, and that human facility in working with structural groupings and symmetric relations develops only in the latter stages of childhood.

These factors provide strong argument that one of the major needs of image understanding systems now is a robust, generic method for representing and processing both group-oriented and symmetry-oriented structural properties of images. This paper illustrates a method for representing the low-level (grouping) structure of segmented images. The structuring process is thoroughly based on an implementation of the factors which a dominant role in human perceptual grouping processes: similarity, proximity, containment, and similar directionality.

The implementation scheme has been applied to both natural (FLIR) and artificial (Bongard) images. The resulting Hierarchal Relational Structures (HRS) provide an organizing schema for grouping related regions for further processing. The HRS further enables a potentially simpler form of representation for higher-order relationships (such as symmetry), and provides a succinct structure to which meaning or object/feature identification processes can be applied.

TABLE OF CONTENTS

1.0 INTRODUCTION: THE NATURE OF THE PROBLEM

The ultimate goal of machine visual perception is to create a complete, multiaspect symbolic representation of the "world" corresponding to the image. This capability would then lead to further capabilities, e.g., an ability to "image" the same scenario from other perspectives, or by manupulating the internal symbolic representation, to "image" potential changes in the scene, as shown in Figure 1.1. These desired capabilities for machine perception correspond to known human capabilities in perception and imagery.

This paper presents an approach to improving machine perception capabilites, or machine image understanding (IU), by drawing substantially from known aspects of human perceptual processes. First, though, it is worthwhile to briefly note the status of machine IU performance. Currently, IU systems can produce "segmented images" from raw image data. [See,e.g., Ballard and Brown, 1982; Marr, 1982]. These segmented images are usually one of two sorts: those which show all the lines or edges which can be found in the original image, and those which show all contiguous regions. Methods which combine these approaches are also available. [E.g., Milgram, 1979] For the purpose of demonstrating the concept, we will use region-segmented images, but the problems and methods will be applicable to either type of segmented image.

Once a region-segmented image is obtained, IU systems tend to characterize each of the segmented regions as much as possible [Panda, 1978; Ballard and Brown, 1982]. Typically this involves obtaining values for region area and extent, describing the shape, color, or texture of the region, and in other ways characterizing the attributes or content of the region. Recent work has emphasized methods which would yield depth or orientation characteristics of the region [Ballard and Brown, 1981; Barrow and Tenenbaum, 1981].

Each of the three major stages discussed above; the original image, the segmented image, and the more symbolically represented 2-D or 2-1/2-D surface knowledge (from the characterized regions) [Marr, 1976], comprises a major form of knowledge representation about the original scene (from which the initial image was taken) [Barrow and Tenenbaum, 1981]. Minimally, two more levels of knowledge representation are desired. The characterized surfaces or regions described above need to be represented in a different internal form which clearly makes explicity the 3-D nature of objects, features, and the background or terrain. In many cases, this will involve a form of geometric 3-D object-centered representation, where each object can be uniquely represented with full symbolic description of its shape. (Excellent work has been done along these lines by Brooks et al, and by other groups of investigators). [Brooks, Greiner, and Binford, 1979; Brooks, 1981; Brooks, 1983]. Other representation schemas may be used for background, terrain, and/or extended terrain features [Sedgwick, 1983].

This does not complete the process of image understanding. While a 3-D geometric representation of a perceived scene is necessary, it must be supplemented greatly by knowledge about the identification and meaning of each object, feature, and aspect of terrain. At this level, objects would be represented in terms of function rather than form. Functionally-related object classes would exist, as would (potentially) schemas for events or object configurations [Tsotos, 1984; Havens and Mackworth, 1983].

Thus, the traditional approach to conceptually designing an image understanding system uses the idea of multiple knowledge representation levels, hierarchically arranged (as shown in Fig. 1.2). The major levels of knowledge representation are (roughly): original image, segmented image, symbolic representation and characterization of the segmented image, (including depth and orientaton information), 3-D geometric form-based representation, and a topmost "meaning" or "interpretation" level. Each of these levels, of course, can obtain several different sub-levels.

There are several major bottlenecks which lie between the current state-of-the-art in machine IU and the desired capabilities for machine IU. These bottlenecks appear to be strongly associated with major transitions in the levels of representation (shown in Fig. 1.2).

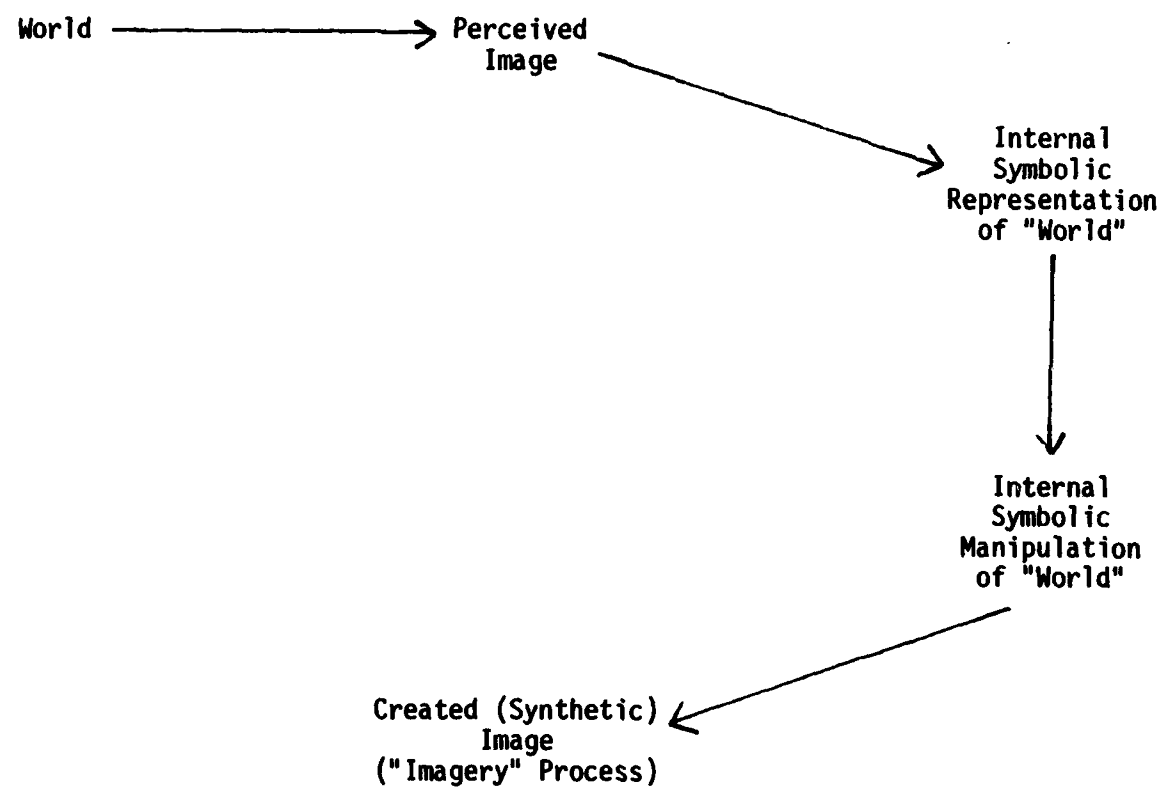


Figure 1.1 Complete, multiaspect, symbolic representation of a perceived image can lead to the ability to create perform "imagery" associated with manipulating the symbolic representation.

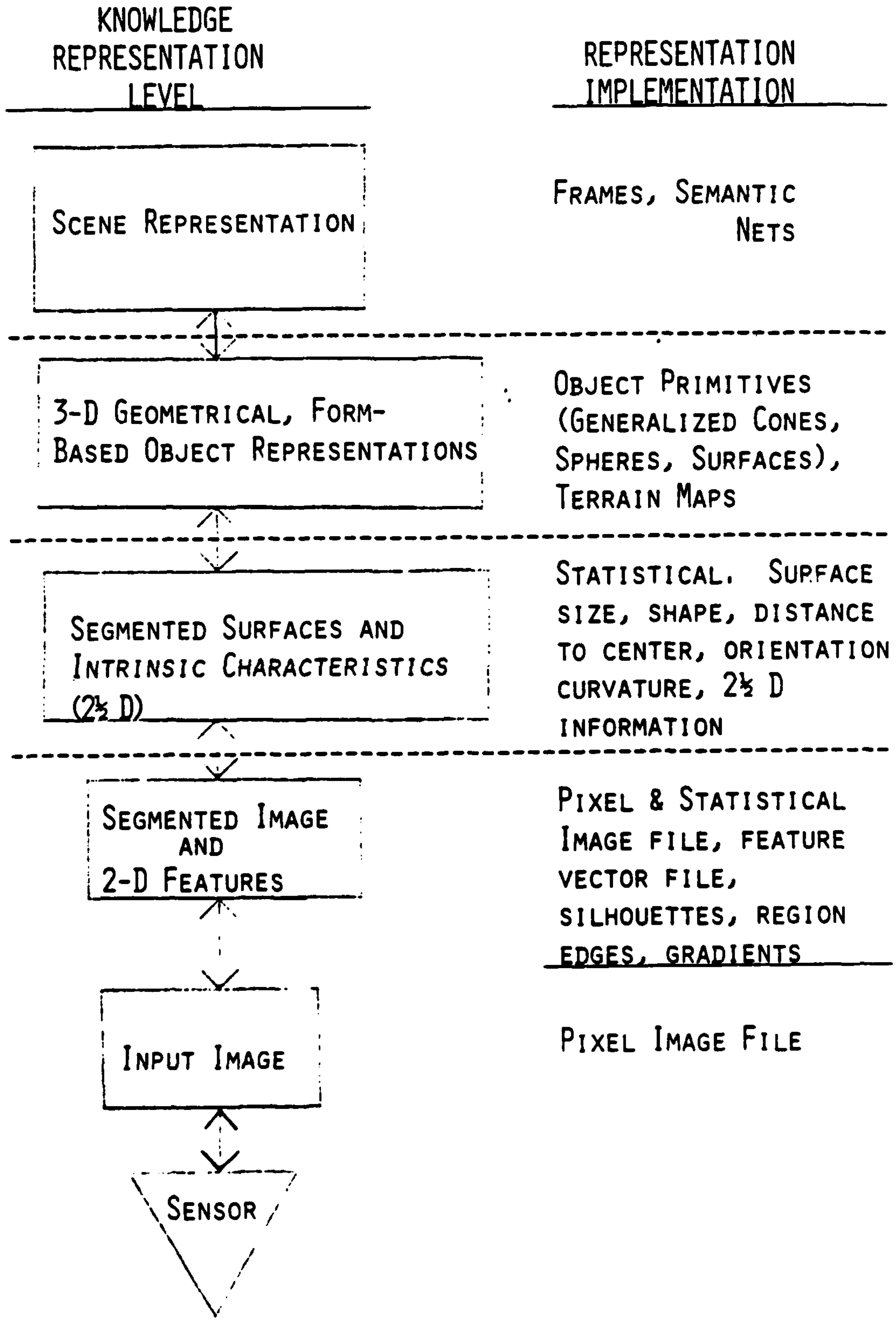


Figure 1.2 Traditional machine IU systems are strongly hierarchical. Each hierarchical level corresponds to a major representation formalism for knowledge associated with the image.

There has been, in recent years, an enormous amount of work aimed at facilitating the transition between these representation levels. Interesting and useful approaches have been suggested, each with greater or lessor degrees of applicability to generic IU systems [Brooks, Greiner, and Binford, 1979; Brooks, 1981, 1983; Binford, 1982 (and references contained therein); Oshima, 1983].

In most recent years, effort has focused on improved characterization of depth or otientation which can be obtained from many sources (shape from texture [Kander, 1979; Stevens, 1981; Beck, Prazalny, and Rosenfeld, 1983], shape from shading [Barrow and Tenenbaum, 1981; Smith, 1983], and other approaches [Stevens, 1981; Haber, 1983]. However, an awareness is emerging that these techniques, no matter how refined and sophisticated they become, will in and of themselves be insufficient to enable inter-level transitions, particularly the crucial transition from symbolic 2-D to 3-D representation [Witkin and Tenenbaum, 1983; Lowe and Binford, 1981, Lowe, 1984].

A major theme of this paper (and the theme of some recent efforts by other *investigators) is that the previous approaches* place a heavy reliance on extracting characteristics (content) of regions in an image. This paper defines an approach to image understanding which incorporate more fully the relationships between the regions, and defines a higher-level concept: that of structure in the image. The "structure" which will be the subject of this paper may be preliminarily defined as "the interrelation of parts as dominated by the general character of the whole" [Webster's Third New International Dictionary].

The goal of this paper is to examine some aspects of the role of structure in human image understanding, and to extract from our knowledge of human visual processes certain capabilities which could potentially play a strong role in machine IU. It also shows how these capabilities could influence the overall performance of machine IU, and illustrates the proposed use of structure in machine visual perception for both artificial and real images.

2.0 THE ROLE OF STRUCTURE IN HUMAN VISUAL PERCEPTION.

A premise for this section is that humans perceive and describe objects and scenes in terms of four major aspects: content, context, structure, and meaning. This simplified premise is offered primarily as an organizing schemata so that progress and capabilities of machine image understanding can be compared with human abilities.

Briefly each of these terms means the following:

- <u>Content</u> - Description of an object (or segmented region in a scene) in terms of its attributes, e.g., shape, texture color, and size.

- <u>Context</u> - Description of the relationships between an object or segmented region) and neighboring objects or regions. Common relationships to consider are the so called "Gestalt" relationships: proximity, similarity, containment (or enclosure), and directionality.

- <u>Structure</u> - Description of an object (or configuration of objects or regions) in terms of the organization which describes how, overall, the interrelations of the parts contribute to perception of the whole.

- <u>Meaning</u> - Naming, interpretation, and/or connotation of a perceived object or region, or "configuration" of objects or regions. Meaning can be related to the specificity of naming, or to the connotated properties of the named object.

A substantial body research has shown that humans work towards and create an internal symbolic representation of objects or scenes. This internal representation contains content, context, structure, and meaning for perceived and imaged scenes.

Although it is by no means clear that humans operate with the same hierarchical representation structure as has been proposed for machine IU systems, it is well known that humans have the capability to symbolically represent and interpret different levels or forms of visually-related knowledge:

1. Representation of 2-D charactersitics of a perceived object or scene. This is most obvious in the case of abstract line drawings or figures where no three-dimensional information is intended. [Vernon, 1953; Kohler, 1947; Koffka, 1935]

2. Internal representation of the three-dimensional nature of a perceived object or scene. Our ability to mentally manipulate such internal three-dimensional models has been amply demonstrated, most notably by the experiments of Shepard [Shepard, 1971].

3. Representation of classes of objects and schemas of events. Human facility for organizing objects and events is well known, and has formed a basis for significant research in artificial intelligence and cognitive science [See, e.g., Minsky, 1975].

These human capabilities evolve over the course of childhood. Work by Binet and Simon showed that the evolution of scene understanding abilities in children developed with age [Binet, 1916]. At three years of age, children could enumerate the objects in a scene. By the age of seven, they could describe objects. Between the ages of seven and fifteen, they developed the ability to describe events and the the relationships between objects and persons. Generally, by the age of eleven, they were able to interpret the picture as a whole [Binet, 1916; Vernon, 1953].

There is some similarity between the evolution of abilities in children and the process by which adults identify a figure. Adults first have a vague "feeling of something", followed by a vague impression of some indefinite object. This is followed by the "generic object stage", at which certain parts of the object stand out more clearly. The next stage (the "specific object stage") is one in which the observer perceives an organization of the parts in the object or figure, while the background fades out. In the last stage, naming of the object occurs [Vernon, 1953].

These stages, and the evolution of perceptual abilities described above, could be approximately described as a procession of capabilities corresponding to the four aspects of perception given earlier: content, context, structure, and meaning.

Early work by researchers in perception showed that the tendency to group objects or regions based on laws of Pragnanz (similarity, proximity, containment, and directionality) played an extremely important role in human perception [Koffka, 1935; Kohler, 1947; Beck, 1966, 1972; Olson & Attreave, 1970; Rock & Brosgole, 1964]. These studies were later advanced to show that humans organized perception so as to create an object or configuration based representation that yielded maximal symmetry under group-theoretic symmetry operators [Garner, 1970; Palmer, 1983 (and references contained therein)].

The perception of structure in abstract figure drawings is an ability which increases with age in children. At the age of three and one half to four years, children can recognize and reproduce the difference between open and closed figures, and from the age of four they begin to recognize the difference in rectilinear and curved figures, and to differentiate among such figure types as squares and triangles. At this age, they also begin to be able to understand object relations [Piaget and Inhelder, 1948].

Young children tend to copy exact details of a complex line drawing, juxtaposed without any idea of the pattern as a whole. They have no concept of the relationships among the details or between details and overall structure. By the age of nine to ten years, children are able to identify the main outlines of structure in an abstract line drawing. It is interesting that naming and verbal analysis of structure appear very involved in the child's ability to extract structural components. By the age of eleven to twelve years, children are able to perceive main outlines of structure, supporting or subsidiary interrelations, and are able to integrate details in a manner corresponding to the original image [Vernon, 1953].

To what extent is structure an important factor in human visual perception? Many indications imply that structure plays a very important role. During this discussion, it will be necessary to distinguish between two levels of structure, will be referred to as "low-level" and "high-level" structure. Low-level structure is the organization imposed on a figure or scene by grouping related objects or regions together according to the Laws of Pragnanz [Kohler, 1947]. High-level structure will be a reworking of the groupings found by low-level structure as to perceive the maximal symmetry of a configuration [Palmer, 1983]. This symmetry may be either translational, regarding a repeated pattern or figure, or it may involve rotation or reflection about a point or axis.

There is some evidence that the role of structure is to aid in the human capability to form internal symbolic representatons of images. This stored internal symbolic representation is what would allow us to recreate scenes once they have been viewed, or to create new or similar scenes [See, e.g.; Shepard, 1978]. An illustration of this is that humans tend to draw a simplified version of a complex, abstract line drawing which they are asked to reproduce from memory. The line drawing produced is not only simpler than the original, it also has a higher degree of symmetry than the original drawing [Vernon, 1953; Shepard, 1978].

Recent research has shown that the right hemisphere of the human brain is highly specialized (in right-handed persons) to handle a wide variety of spatial relational processing. Damage to this hemisphere can result in severely impaired capabilities to interpret and understand imagery. Overall, research in localization of brain functioning supports the view that relationships between regions, and information which can be drawn from these relationships (structure) is fundamental to human visual perception [Kimura and Durford, 1974; Levy et al, 1983].

3.0 THE ROLE OF STRUCTURE IN MACHINE VISUAL PERCEPTION

As stated in the Introduction section, the goal of machine visual perception is to construct a complete, accurate, multi-aspect symbolic representaton of the "world" perceived through the imaging sensor. This would enable an intelligent machine to perform further operations (which might be somewhat akin to imagery). In order to be able to construct a useful internal symbolic representation, the machine would have to have image understanding capabilities similar to those described earlier for humans. In the context of machine image understanding these are:

o Content - Segmentation of the input image into appropriate regions, and characterization of these regions in terms of size, shape, color, texture, and other useful attributes.

o Context - Internal representation of the simple interrelationships between regions. Important relations to take into account would include similarity, proximity, containment (or enclosure), and directionality.

o Structure - Two levels of structure are appropriate even at the 2-D representation level. These are:

- Low-level structure - Grouping "related regions" so that they may symbolically be considered as forming parts of a "whole".

- High-level structure - Determine the group-theoretic operators which may describe a configuration or pattern expressed in single regions or groups of regions. This may involve identifying points or axes of symmetry, or translational invariance for repeated forms. This structure must be robust enough to account for effects induced by perspective.

o Meaning - Recognize, classify, and identify important objects and features. Associate contextual information with certain object classes (e.g., capabilities of or uses for certain objects). Interpret events, or "meaningful" configurations of objects or features.

These four desired capabilities of image understanding systems are geared towards enabling the system to contruct an internal symbolic representation of the world corresponding to the perceived image. It is still likely that at least three major levels of representation will be necessary. These three levels would be:

o 2-D Symbolic - Symbolic representation of the segmented regions and their attributes (content), their relationships (context), and both low- and high-level structures describing the configurations of the regions.

o 3-D Form-Based - Representation of the probable three-dimensional forms which yield the image obtained by the sensor. This can include object-centered representations for segmented objects, and terrain-based information for representation of the background or terrain.

o Object/Feature Classes - Representation of all identifying and connotative information associated with the perceived objects, features, and terrain.

In comparison with the desired capabilities of machine visual perception systems, existing systems have exceedingly limited capabilities. Current machine vision systems focus on object identification where the domain of objects under consideration is relatively small [Binford, 1982; Rosenfeld, 1983]. These systems rely to a great extent on object attributes, such as size, shape, intensity or color, texture, and velocity (for real time vision systems). Some systems make use of relationships between different regions or edges of an object. In these cases, the type of relational information extracted is very limited, and does not inlcude the full range of information that comprises the domain of the Laws of Pragnanz. The relationships considered also tend to be highly specific to the object domain, and lack the generality of human perceptual groupings [Brooks, 1981, 1983; Oshima, 1983; Shirai, 1978]. No active systems take into account the full nature of structure in constructing internal symbolic representations of the perceived image, although recent work by Lowe represents a significant step in this direction [Lowe and Binford, 1981; Lowe, 1984]. Earlier work in visual analogic relations is not strong enough to form a substantial base for image understanding [Evans, 1968].

An alternative organization for a machine image understanding system is shown in Fig. 3.1. While this proposed system design maintains the fundamentally hierarchical approach of early designs for IU systems [Barrow and Tenenbaum, 1981], it departs radically from earlier systems through inclusion of context (relationships), and structure at the crucial 2-D (or 2-1/2D) and 3-D symbolic levels. In this sense, there is somewhat of an analogy to the lateralization of visual processes in the human brain.

In the 2-1/2-D level, content information leads to context, from which low-level structure can be gained. Structure provides groupings of regions which facilitate object/feature characterization and identification. Using both 2-1/2-D content and structure information, a fuller interpretation of the 3-D characteristics of an object or feature is possible. At the 3-D level also, the contextual knowledge is facilitated by both 2-1/2-D context and structure. Finally, the 2-1/2-D structure leads into 3-D structure, which will include full 3-D symmetry understanding.

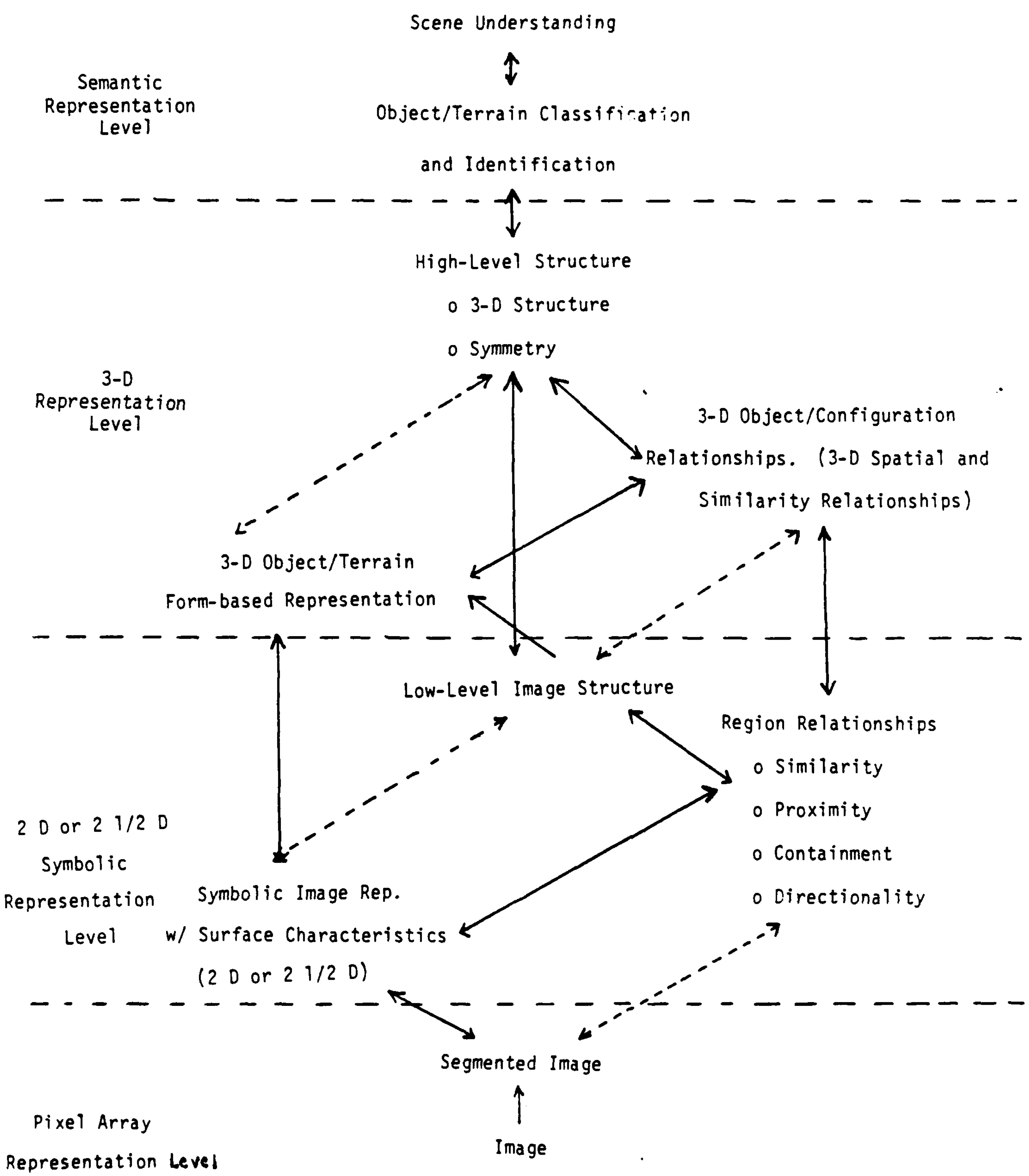


Figure 3.1 Conceptual design (in terms of representation levels) for a machine IU system which would incorporate use of relationships between segmented regions, and image structure based on the relationships between regions.

4.0 CAPABILITIES ENABLED THROUGH CONSIDERATION OF STRUCTURE IN IMAGES

The structure of regions in an image plays an important role in enabling perceptual processes and image understanding. By representing the structure of segmented regions, as well as the attributes of and relationships between these regions, a new form of information is made available to the knowledge bases and inference engines which would operate on the image. For the remainder of this paper, the emphasis will be on the representation of low-level (grouping) structure for images, and on the knowledge which can be gained from this structural representation through inference.

In brief, the mechanism for evolving a representation of low-level (Gestalt grouping) structure is as follows: Each segmented region in an image is symbolically represented by a "node", which will have appended to it content and context information. A parameterized form of the Gestalt "Laws of Pragnanz" is used to determine the relationships between nodes (contextual information). (These relationships are detailed in Table 4.1.) This allows the most "closely-related" nodes to be identified and grouped together. A "cluster node" symbolically represents this grouping of closely-related nodes. Content and context information for the new cluster node is calculated. The process is repeated, creating higher levels of cluster nodes, until the entire image (or segmented object) is represented by only a few cluster nodes. Decomposing the hierarchical cluster-node structure allows access to the components of each group of nodes. The result of this process is that regions which would be grouped together by our human perceptual processes are similarly grouped together by the structural process (referred to as Hierarchical Region Structure (HRS)).

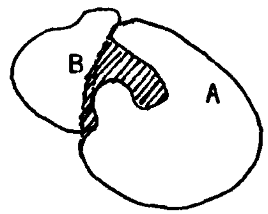
An example is shown in Fig. 4.1. Fig. 4.1(a) shows a Forward-Looking InfraRed (FLIR) image of a road. We would naturally group together the two semi-parallel strips together as a structural unit, or "configuration". Moreover, we would naturally consider the left and right (bright) sides bordering the road each to be "units" (in a symbolic sense) despite the fact that there are three regions

TABLE 4.1

RELATIONSHIPS BETWEEN REGION PAIRS

RELATIONSHIP	DESCRIPTION	MATHEMATICAL EXPRESSION
Size Similarity	$R_{size}(A,B)$ measures the similarity in size between two regions. $R_{size}=1$ for regions of the same size, and decreases exponentially with difference in size.	$R_{size}(A,B)=\exp[P_{size} * ABS[Size(A)-Size(B)]]$ where Size(A), Size(B) are the areas (in pixels) of regions A and B, respectively.*
Intensity Similarity	$R_{int}(A,B)=1$ for regions of the same intensity, and decreases exponentially with difference in intensity.	$R_{Int}(A,B)=\exp[P_{int} * ABS[Int(A)-Int(B)]]$ where Int(A), Int(B) are the intensities of regions A and B, respectively.
Shape Similarity	Regions are classified into one of three different shapes (Convex, Elongated, Irregular).	$R_{shape}(A,B)=1$ if $Shape(A)=Shape(B)$, $=0$ if $Shape(A)\neq Shape(B)$ where Shape may be convex, elongated, or irregular.
Texture Similarity	Segmentation method determines the number of different texture classes.	$R_{tex}=1$ if $Tex(A)=Tex(B)$, $=0$ if $Tex(A)\neq Tex(B)$ where Tex(A), Tex(B) are the texture classes for regions A and B. Algorithm for determining texture classes under development.

* P_i is a scaling parameter associated with each function R_i.

Proximity	Region proximity is calculated for two cases: Case 1: Disjoint regions (No boundary points in common. As first-cut approach, R_{prox} is calculated using distance between region boundaries. Later, distance between region centroids will be included to account for the fact that measures of proximity of disjoint regions must account for both the size of the regions and the separation distance between regions. $0<R_{prox}<1$.	Case 1: Disjoint regions. $R_{Dprox}^{(A,B)}=\exp\{-P_{Dprox} \times [\mathrm{EUCDIST}\ (\mathrm{Bound}(A), \mathrm{Bound}(B))]\}$ where EUCDIST (Bound(A), Bound(B)) is the Euclidian distance between the boundary pixels of regions A and B, where the boundary points lie along the line connecting the centroids. EUCDIST (Bound(A), Bound(B)) 
	Case 2: Adjacent regions (Regions share one or more boundary pixel.) $R_{prox}^{(A,B)}$ measures the degree to which region A is proximal to region B. (This can be different than the value for $R_{prox}^{(B,A)}$.) $1 \leq R_{prox}$.	Case 2: Adjacent Regions $R_{Aprox}^{(A,B)}=1 + P_{Aprox} \times \dfrac{\mathrm{NUMBOUND(A)} \cdot \mathrm{NUMBOUND(B)}}{\mathrm{NUMBOUND(A)}}$ where NUMBOUND(A), NUMBOUND(B) are the number of boundary points on regions A and B respectively, and the numerator is the number of boundary points which border both regions A and B.
Containment	$R_{cont}^{(A,B)}$ is the measure of how much region A contains region B. It is calculated based on how much of region B is inside the convex hull of region A. (The convex hull of A is the closest convex curve that can be fit around A.)	$R_{cont}^{(A,B)}=\dfrac{\mathrm{CONTAIN(A,B)}}{\mathrm{SIZE(B)}}$, where CONTAIN(A,B) is the number of pixels of B inside in the convex hull around A.  Hatched area is the part of B contained by convex hull of A region A.

Orientation	$R_{orient}^{(A,B)}$ measures the degree to which two elongated regions are oriented in the same direction. For regions oriented in exactly the same direction, $R_{orient}^{(A,B)}=1$.	$R_{orient}^{(A,B)}=\exp\{-P_{(orient)} \times \frac{\theta}{180}\}$ where θ is the angle (in degrees) between the median axes of two elongated regions.
Directionality	$R_{dir}^{(A,B)}$ measures the degree to which two moving regions are moving in the same direction and with the same velocity. If two regions have exactly the same direction and velocity, $R_{dir}^{(A,B)}=2$.	$R_{dir}^{(A,B)}=\exp\{-P_{dir} \times \frac{\theta}{180}\}$ $+ \exp\{-P_{vel} \times ABS(VEL(A)-VEL(B))\}$, where θ is the angle (in degrees) between the direction of motion of moving regions A and B, and VEL(A) and VEL(B) are the velocities of regions A and B, respectively.
Colinearity	$R_{col}^{(A,B)}$ measures the degree to which two elongated regions have medial axes which are colinear.	Algorithm under development.

Figure 4.1 (a) The original FLIR image of a road which will be used to demonstrate application of the Hierarchical Region Structure method.

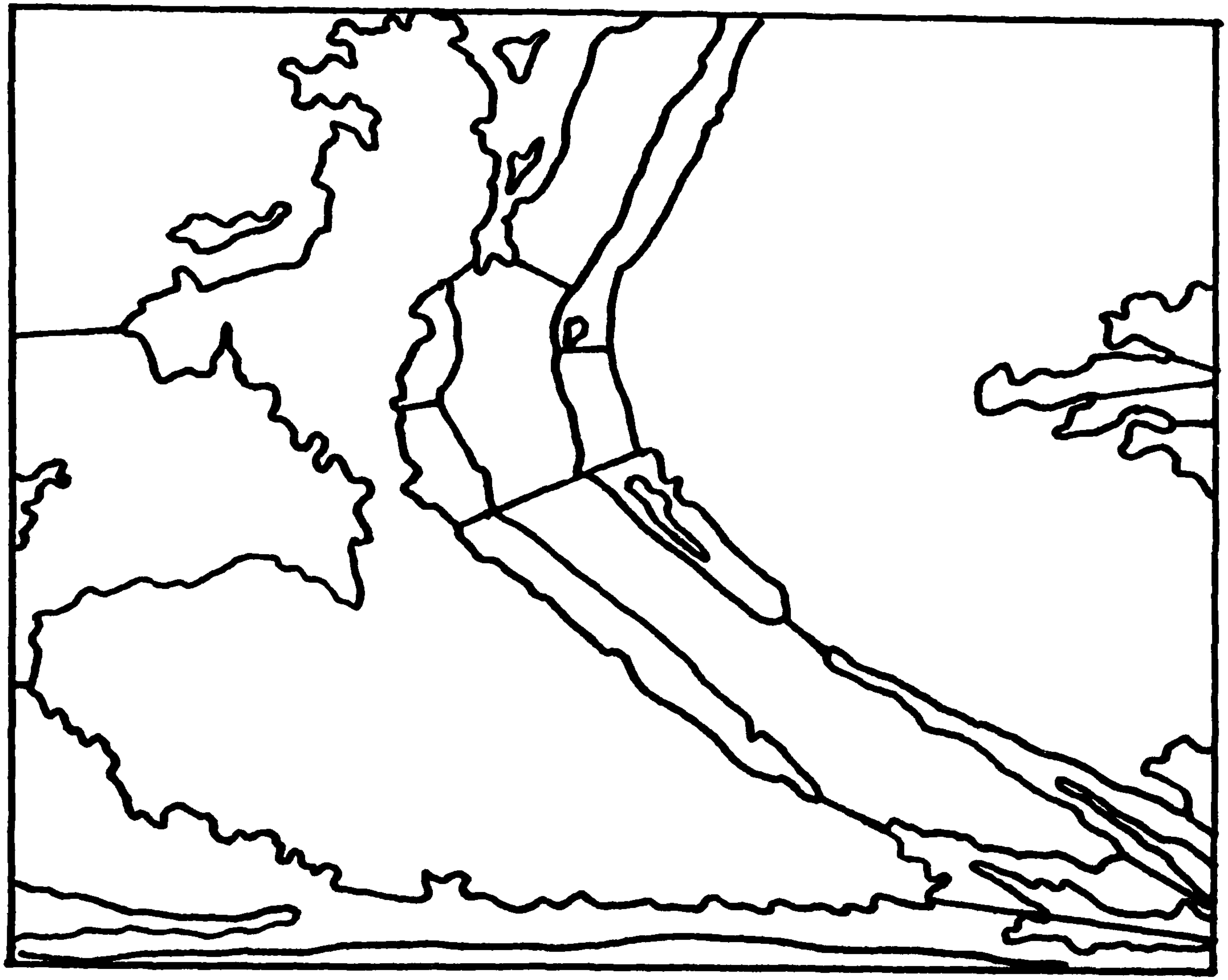


Figure 4.1 (b) A simplified segmented image of the FLIR road image shown in Figure 4.1 (a). Some manual resegmentation has been done on this segmented image.

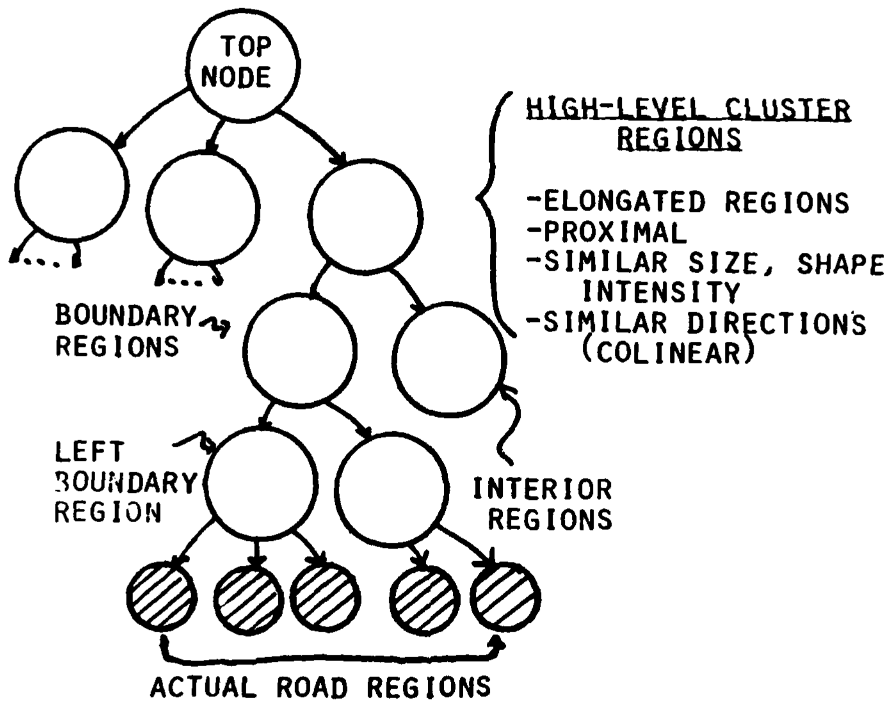


Figure 4.1 (c) The Hierarchical Region Structure of the segmented road regions shown in Figure 4.1 (b).

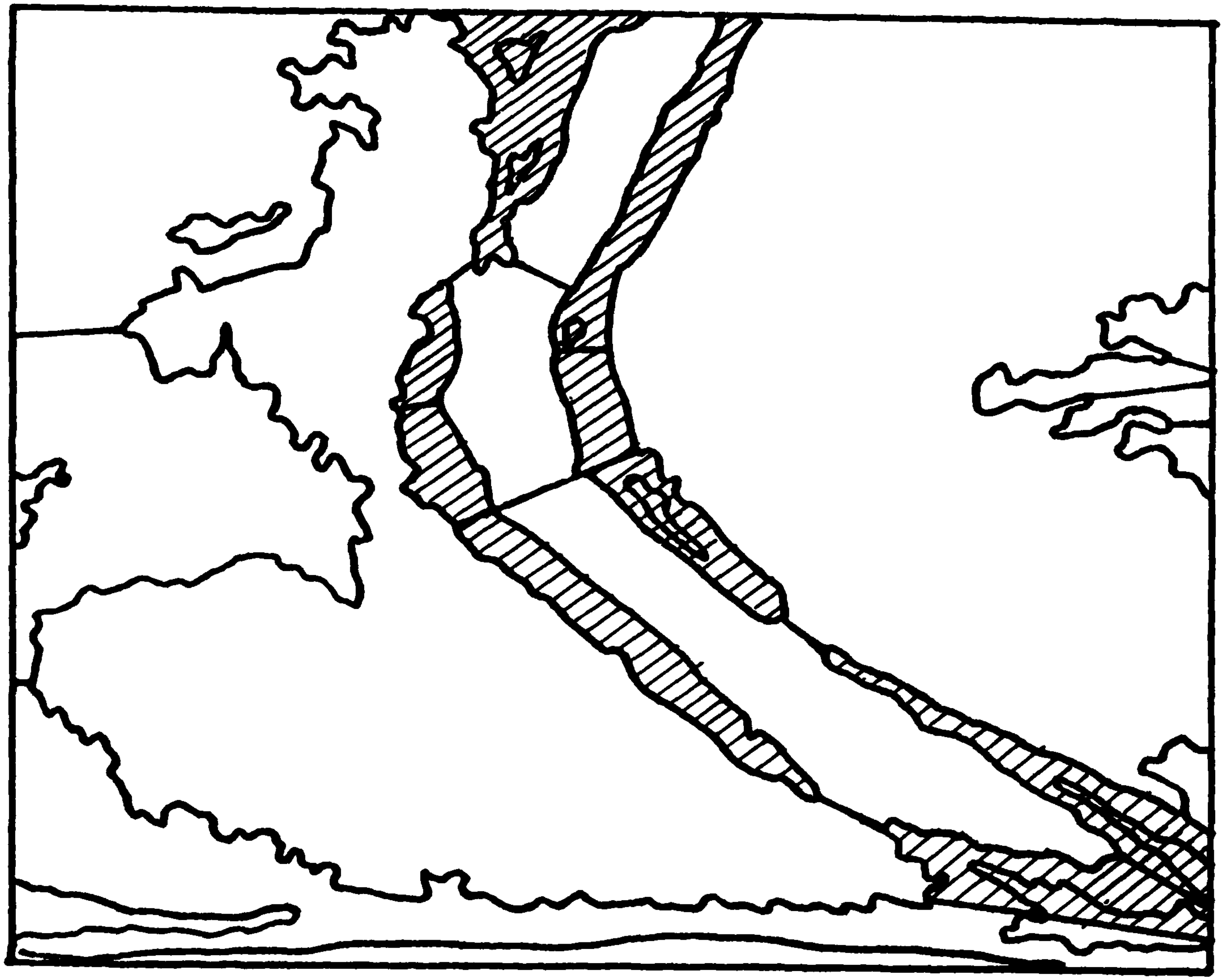


Figure 4.1 (d) The hatched road regions correspond to the hatched lower-level circles in Figure 4.1 (c). They are symbolically represented by a single node at a higher level in the HRS shown in Figure 4.1 (c).

corresponding to the left-hand road border, and two regions on the right. Current image processing algorithms can segment out these regions (to greater or lesser degrees of satisfaction, depending on the methods used), and characterize each of the segmented regions. However, a mechanism to relate the segmented road regions together into a single structural whole is not a part of generic IU technology. In this sense, the grouping offered by the HRS approach represents a strong development in IU capabilities.

Fig. 4.1 (b) shows a segmented image based on Fig. 4.1 (a). Fig. 4.1(c) shows the part of the Hierarchical Region Structure which relates to the road area. The hatched nodes at the bottom of Fig. 4.1(c) correspond to specific regions in the segmented image, shown hatched in Fig. 4.1(d). The three left-side border road regions are grouped together (using high-value relations of similarity , proximity, and similar direction), as are the two right-side border regions. These form the two cluster nodes which can be seen at the second level from the bottom of Fig. 4.1(c). These two nodes are grouped together on the basis of their similarity and similar directionality, and form a new cluster node (third level from bottom). This node joins with a cluster node representing the interior of the region to form an overall road node (one level down from the top). Similar processes would group together the regions on the right and left sides of the region.

Using this hierarchical structure, to which content and context information would be appended, an inference engine could search among the top layers of the structure for a cluster node which has characteristics and internal relationships (context) corresponding to the known content and context of a road. A strong advantage of this approach is that at this point, all of the road regions would be symbolically linked together and could be treated as a unit, thus enabling further high level processing.

Figs. 4.2 - 4.4 further illustrate the capabilities offered by a representation of low-level structure in an image, as well as the limitations inherent in this approach. These figures are all based on Figs. 88, 89 and 90 in Bongard's one hundred "Problems for the Recognition Program" [Bongard, 1970]. Each of the original problems consisted of two sets of six small figures each. The goal in each case was to determine what unique characteristic described one set of figures which differentiated it from the other set of figures in the same problem. For succinctness, I have used only three of the figures of each of the sets, instead of the original six. Each of the original figures used for discussion here is shown in the left-hand column, and a hierarchical structural representation for it is shown on the right.

In the structural representation, each node which is a termination of the downward-pointing tree-like structure represents one of the white or black ellipses in the original figure. (The notations of "W" for white and "B" for black are added underneath the terminal nodes of the hierarchical structure in order to facilitate comparison with the original.)

The figures in the first set of Fig. 4.2 each have three ellipses, those in the second set of the same figure each have five ellipses. In this case, the addition of an hierarchal structure which describes the structural groupings of each of the figures does little to help. Differentiation between the two sets can be based on content alone. Such differentiation is well within the province of traditional image understanding systems, which could extract the content of each figure, and infer the nature of the set definitions: three or five ellipses per set, respectively.

Fig. 4.3 presents a slightly more complex situation. In this case, the content of each figure is not sufficient to fully differentiate the sets. Some figures have three ellipses, others five, others as much as fifteen. The key feature here is the grouping of the ellipses together. This is shown vividly in the accompanying hierarchical structures. Each of the structures for the first set has three major

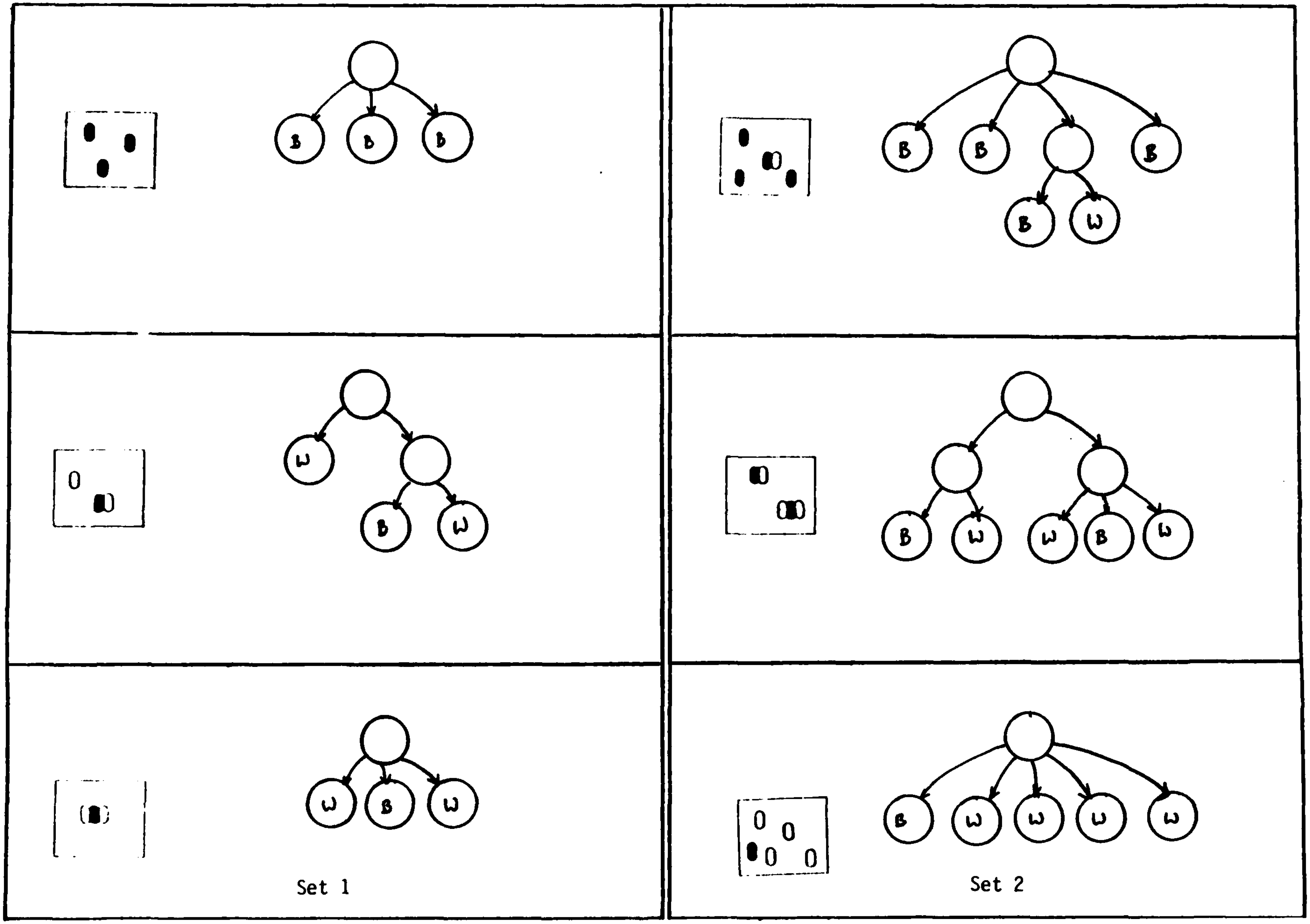


Figure 4.2 A visual analogy problem (Bongard, Problem 88) in which structure is not necessary to complete the analogy. (Set 1), the first set of three figures (enclosed in boxes), including HRS representations of the structure of those three figures; (Set 2), the second set of three figures, including HRS representations of the second set.

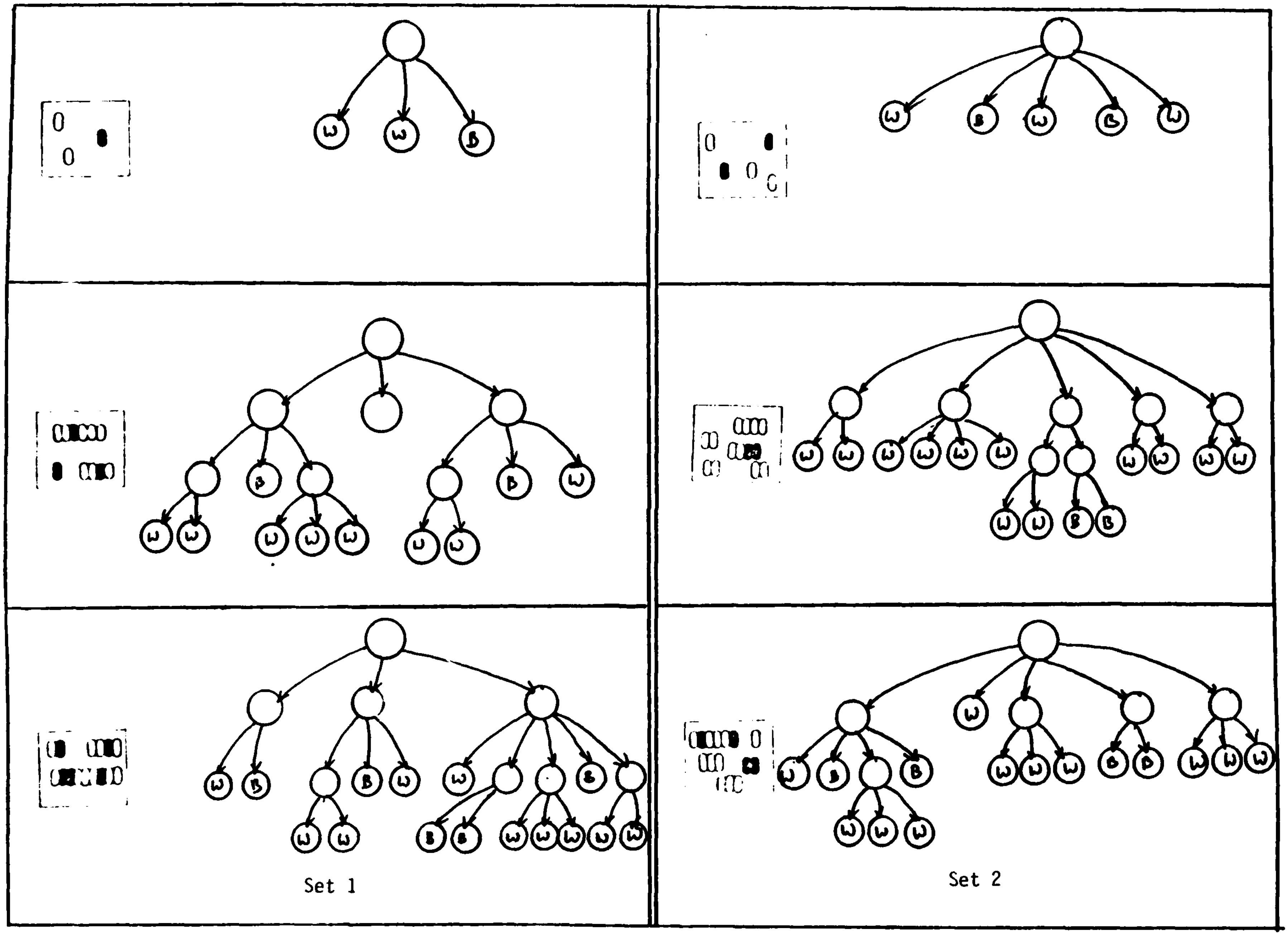


Figure 4.3 A visual analogy problem (Bongard, Problem 89), in which a low-level structural representation is necessary for a machine to solve the analogy.

"cluster nodes" (in an expanded use of the term, since sometimes a "cluster node" will correspond to a single node of the original image). Each of the structures in the second set has five cluster nodes. In each case, the structures are built using heuristics and processes based on the Gestalt Laws of Pragnanz. For simple figures such as these, the resulting structures are very strongly determined, with little ambiguity in terms of the groups which will be formed. The information produced by this Gestalt-like grouping or structuring process presents a major step beyond that provided by traditional IU systems.

Despite the clear advantage that such structural groupings offer, it is important to be clear on the limitations of such information. It is necessary to understand that the information offered by the low-level structural process is devoid of both meaning and representation of symmetry. These limitations are illustrated by Fig. 4.4. Here, although the hierarchical structure clearly groups together the adjacent white ellipses before clustering them with the black ellipses, it is by no means clear that this is sufficient to distinguish between the two sets. The distinguishing criterion in this case (three groups of white ellipses in the first set, four in the other), would have to be made at a higher inferential level than that of the previous figure. Thus, while low-level structuring offers a substantial body of information in a form which enables rapid processing, it is only another step in developing mature IU systems; it is not by itself sufficient to solve all IU problems. Nevertheless, the advantages offered by the use of low-level structure in IU systems is so strong that this area warrants active attention.

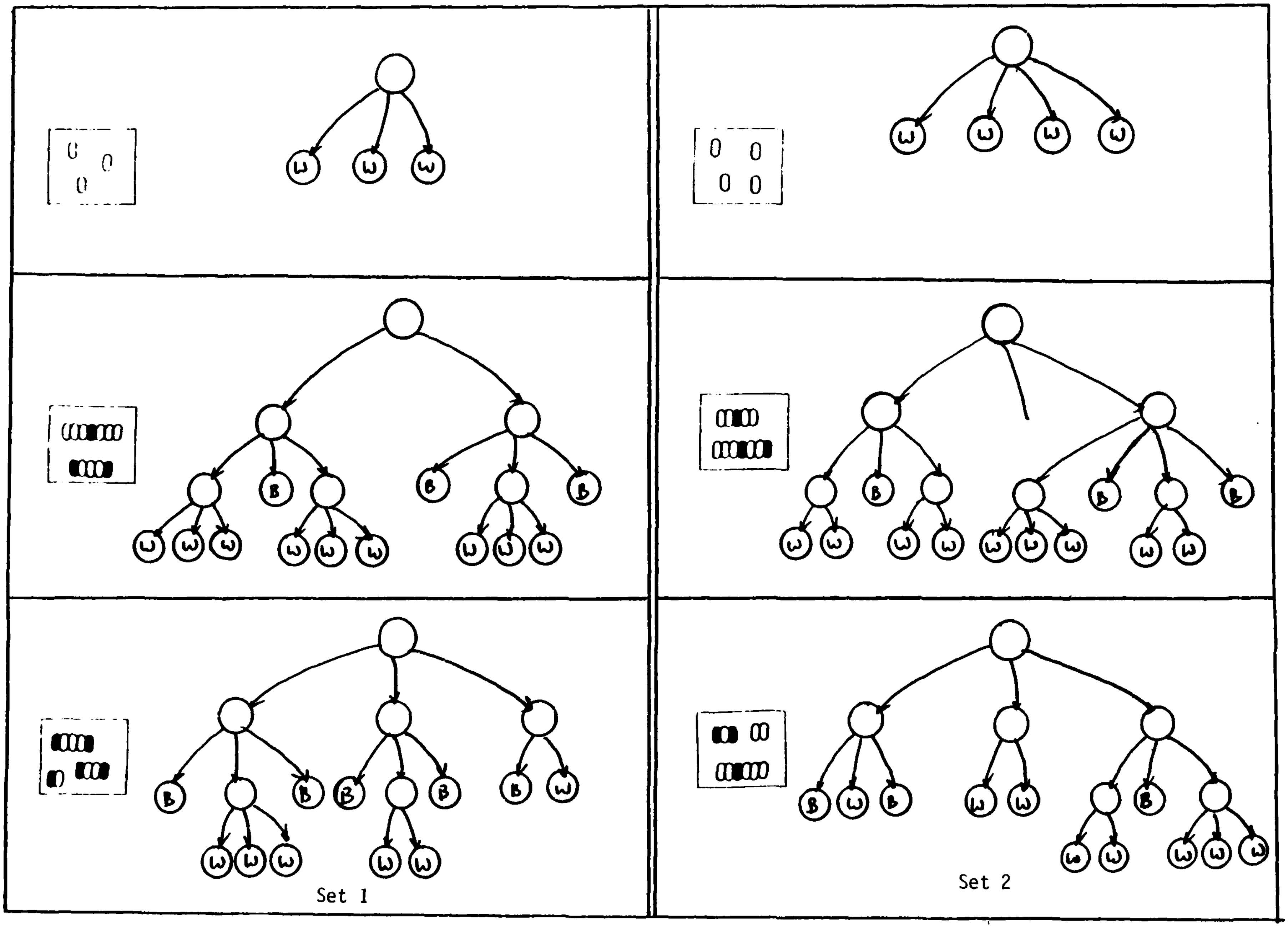


Figure 4.4 A visual analogy problem (Bongard, Problem 90), in which low-level structure is, by itself, insufficient to readily enable solution of the analogy.

5.0 CONCLUSION: ASSESSMENT AND IMPLICATIONS

For some time, there has been a growing awareness among the image understanding resear munity that the traditional approaches were not yielding satisfactory results in terms of desired performance. These traditional approaches have been heavily weighted towards obtaining the characteristics of segmented regions. Even the more recent work, leading potentially to realization of Marr's 2-1/2-D primal sketch, has simply focused on obtaining depth or orientation cues about regions, which is an elaboration of the characterization process.

In contrast to this approach, researchers in human perceptual processes have long been aware of the importance of the use of relationships between regions, leading to grouping of regions. This capability to form groups of highly related regions is a fundamental (low-level) form of structuring the information in an image. Further work showed that humans make extensive use of the symmetry properties of image configurations in building up internal symbolic representations of the perceived images. It has become apparent that this capability is not a trivial one, and that human ability in working with structural groupings and symmetric relations develops only in the latter stages of childhood.

These factors provide strong argument that one of the major needs of image understanding systems now is a robust, generic method for representing and processing both group-oriented and symmetry-oriented structural properties of images. This paper illustrates a method for representing the low-level (grouping) structure of segmented images. The structuring process is thoroughly based on an implementation of the factors which a dominant role in human perceptual grouping processes: similarity, proximity, containment, and similar directionality.

The implementation scheme has been applied to both natural (FLIR) and artificial (Bongard) images. The resulting Hierarchal Relational Structures provide an organizing schema for grouping related regions for further processing. The HRS further enables a potentially simpler form of representation for higher-order relationships (such as symmetry), and provides a succinct structure to which meaning or object/feature identification processes can be applied.

While this paper has brought up the need for representing higher-order (symmetry-based) structure, it has not attempted to demonstrate a form or mechanism for building this representation. The factors which might be most fruitful yields form this work are the realizations that first, image structure representations store a valuable form of image-based information. Second, investigation into human visual processes can continue to yield valuable insights and ideas for generic, machine-implemented image understanding systems. Thirdly, while structure may be rightly regarded as a valuable component in image description, it is not to be confused with interpretation or understanding of the image. In this perspective, the use of structure in machine image understanding needs to be regarded as one among the many possible elements which can ultimately enable full image understanding.

Ballard, D.H., & Brown, C.M. Computer vision. New Jersey: Prentice-Hall, 1982

Barrow, H.G.& Tenenbaum, J.M. Computational Vision, Proc. of IEEE, 1981, 69, 572-595.

Beck, J. Effect of orientation and of shape similarity on perceptual grouping. Perception and Psycophysics, 1966, 1, 300-302.

Beck, J. Similarity grouping and perceptual discriminability under uncertainty. American Journal of Psychology, 1972, 85, 1-19.

Beck, J., Prazdny, K., & Rosenfeld, A. A theory of textural segmentation. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), Human and Machine Vision, New York: Academic, 1983.

Binet, A. The Development of Intelligence in Children. Baltimore: Williams and Wilkins, 1916.

Binford, T.O. Survey of model-based image analysis sytems. The Internat. J. of Robotics Research, 1982, 1(1), 18-64.

Bongard, M. Pattern Recognition, Trans. by T. Cheron, Ed. by J.K. Hawkins, New York: Spartan, 1970

Brooks, R.A., Greiner, R., & Binford, T.O. The ACRONYM model-based system. IJCAI-79, 105-113.

Brooks, R.A. Symbolic Reasoning among 3-D models and 2-D images. Artificial Intell. J., 1981, 17, 285-348.

Brooks, R.A. Model-based three-dimensional interpretations of two-dimensional images. IEEE Transactions PAMI, 1983, (5)(2), 140-150.

Evans, T.G. A program for the solution of a class of geometric-analogy intelligence-test questions. In M. Minsky (Ed.), Semantic Information Processing, Cambridge: MIT Press, 1968.

Garner, W.R. Good patterns have few alternatives. American Scientist, 1970, 58, 34-42.

Havens, W.; Mackworth, A. Representing knowledge of the visual world. Computer, October 1983, 90-96.

Kender, L.R. Shape from texture: A computation paradigm. Proc. DARPA Image Understanding Workshop, April 1979, 134-138.

Kimura, D. & Durnford, M. Normal studies on the function of the right hemisphere in vision. In S.J. Dimond and J.G. Beaumont (Eds.) Hemisphere Function in the Human Brain. New York: Wiley and Sons, 1974.

Kohler, W. Gestalt Psychology. New York: Liveright, 1947.

Koffka, K. Principles of Gestalt Psychology. New York: Harcourt, Brace and Co., 1935.

Levy, J. Heller, W. Banich, M.T. & Burton, L.A. Are variations among right-handed individuals in perceptual asymmetries caused by characteristic arousal differences between hemispheres? J. Experimental Psych: Human Perception and Performance, 1983, 9, 329-359.

Lowe, D.G. & Binford, T.O. Perceptual organization as a basis for visual recognition. Proc. National Conf. on A.I., Aug. 22-26, 1983, Wash. D.C. (AAA1-83), 255-260.

Lowe, D.G. Perceptual Organization and Visual Recognition. Doctoral dissertation, Computer Science Dept., Stanford University, 1984.

Marr, D. Early processing of visual information. Phil. Trans. R. Soc., 1976, B275, 483-524.

Marr, D. Vision. San Francisco: W.H. Freeman and Co., 1982.

Milgram, D.L. Region extraction using convergent evidence. Comp. Graphics and Im. Proc. 1979, 11, 1-112.

Minsky, M. A framework for representing knowledge. In P.H. Winston [Ed.] The Psychology of Computer Vision. New York: McGraw-Hill, 1975.

Olson, R.K., and Altneave, F. What variables produce similarity grouping? American Journal of Psychology, 1970, 83, 1-21.

Oshima, M.; Shirai, Y. Object recognition using three-dimensional information, IEEE Trans. on PAMI, 1983, PAMI-5(4), 353-361.

Palmer, S. The psychology of perceptual organization: A transformational approach. In J. Beck, B. Hope, and A. Rosenfeld, (Eds.) Human and Machine Vision. New York: Academic Press, 1983.

Piaget J. & Inhelder, B. La Representation de l'Espace chez l'Enfant. Paris: Presses Universitaires de France, 1948.

Rock, I. & Brosgole, L. Grouping based on phenomenological proximity. J. of Experimental Psychology. 1964, 67, 531-538.

Rosenfeld, A. Computer vision research for industrial applications. Proc. IEEE Conference on Trends and Applications, Gaithersburg, MD, May 1983, 43-49.

Sedgwick, H.A. Environment centered representation of spatial layout: available visual information from texture and perspective. In Human and Machine Vision. J. Beck, B. Hope, & A. Rosenfeld (Eds.) New York: Academic, 1983.

Shepard, R.N. & Metzler, J. Mental rotation of three-dimensional objects. Science. 1971, 171, 701-703.

Shepard, R.N. Externalization of mental images. In B.S. Randhawa & W.E. Coffman (Eds.), Visual Learning, Thinking, and Communication. New York: Academic, 1978.

Shirai, Y. Recognition of man-made objects using edge cues. In A. Hanson & E. Riseman (Eds.), Computer Vision Systems. New York: Academic, 1978.

Smith, G.B. The relationship between image irradience and surface orientation, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1983, 14-19.

Stevens, K.A., Surface perception from local analysis of texture and contour, MIT AI TR-512, 1979.

Stevens, K.A., The visual interpretation of surface contours. Artificial Intelligence, 1981, 17, 47-73.

Tsotsos, John K., Knowledge and the visual process: Content, form and use. Pattern Rec., 1984, 17(1), 13-27.

Vernon, M., A Further Study of Visual Perception. Cambridge: Cambridge U. Press, 1952.

Witkin, A.P. & Tenenbaum, J.M., On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), Human and Machine Vision, New York: Academic, 1983.

TR85.2

A KNOWLEDGE-BASED IMAGE SEGMENTATION SYSTEM

ABSTRACT

This paper describes a knowledge-based image processing system for the segmentation of outdoor scenes. The system consists of four main units which perform the tasks of: goal determination, preprocessing, segmentation, and region evaluation. The system uses information on the mission goals, sensor characteristics, and data measured from the scene, as well as knowledge about the performance of the individual image processing operators.

The hierarchical image understanding system described here has as its primary goal to segment visual and infrared imagery of outdoor scenes. It performs image or region resegmentation and/or reprocessing by intelligently deciding what image processing operations to use, "measure" the effects on the current image and make a decision as to what the next operation should be. The system has been implemented in a Symbolics 3600 as a collection of production rules acting on a blackboard-type scene knowledge representation called Archival Scene Model. Representative results are described in the paper.

TABLE OF CONTENTS

1.0 INTRODUCTION

One of the fundamental challenges in computer vision is obtaining useful segmentation. In region based segmentation, the goal is to outline regions which accurately correspond to actual structures in the scene. The quality of the segmentation is crucial to effective machine image understanding. Without the capability for good segmentation of physically meaningful regions, even the most intelligent processing cannot achieve satisfactory scene interpretation.

The partitioning of a scene into regions based on additional knowledge, such as range or other ancillary knowledge, can yield very meaningful scene interpretation results [1]. Traditional state-of-the-art image segmentors use some background adaptive technique and threshold in order to meet their requirements. However, such techniques are "blind" in their adaptation and could be severely affected by object size or sensor characteristics instability.

In the past, the most useful segmentation results have been achieved using model driven techniques which were tuned to work for specific situations. These segmentation methods work well for very focused applications such as finding bright targets in low clutter scenes. The performance of these segmentors goes down rapidly, however, when they have to deal with a wider range of imagery. There are many applications where computer vision systems have to function over a wide range of situations. For example, the computer vision system for an autonomous land vehicle must function in a variety of situations which will be affected by different sensors, terrains, weather conditions, and even the time of day. There are currently no segmentation methods that work well over a wide variety of situations.

A way to improve segmentation results over a wide range of scenarios is to use AI techniques to bring more knowledge to bear on the problem. Using AI techniques, such as knowledge on the terrain or weather conditons can guide and control the segmentation process. This knowledge can come from other sensors, the processing results from previous frames, or even from pre-mission training.

Various knowledge driven image processing techniques have been studied by many. Duane et. al. [1985] use knowledge driven production rules to evaluate the region segmentation of an image [2]. This evaluation is used to group smaller regions into larger ones. This evaluation can also resegment a region with new parameters. The segmentation is performed by a single routine that is data driven using no other knowledge. On the other extreme Nazif and Levine [1984] use production rules to control all aspects of the segmentation process [3]. Rules control the analysis and groupings of lines, and regions as well as the scheduling of different segmentation tasks.

Our knowledge based segmentation approach makes use of an in house existing library of image segmentors. The knowledge based control chooses segmentors from this library of operators. It uses external knowledge, and knowledge extracted from the image, to choose the best operator and parameter values for a given situation.

2.0 APPROACH

To use knowledge effectively, the algorithm segmentation process is divided into independent knowledge driven algorithm modules. Each of the algorithm modules performs a specific step in the processing flow. To make a given unit knowledge driven, any knowledge that could possibly aid in the task of the module must be identified and quantified. It is not important where this knowledge comes from, just that it can be known and can be useful in performing the module's task. For example, an algorithm module to do noise cleaning could use knowledge about the overall contrast of the image to determine which noise cleaning operators to use. This knowledge could come from actual measurements of the image or it could be derived from other knowledge such as the type of sensor used, the weather conditions, and the time of day.

All of the information known about a scene is held in a central data based called the global knowledge base (GKB). The GKB handles all of the knowledge transfer between different modules in the system. Each module takes knowledge it needs from the GKB and returns any new knowledge generated to the GKB.

Once the pertinent knowledge has been identified, the operation of each module is designed to be controlled in terms of this knowledge. Production rules have proved to be a good method for knowledge based control within a module. A typical rule is made up of an anticedent and a consequent. The anticedent consists of one or more tests of information in the GKB. If all of the tests in the anticedent are true, then the consequent is executed. The consequent consists of one or more actions that can either change information in the GKB or execute an image processing function. Using production rules makes the control of the segmentation processing knowledge driven, but also makes it flexible and modifiable. Production rules can be added or deleted as new types of knowledge or new image processing routines are added to the system.

3.0 SYSTEM DESCRIPTION

A prototype knowledge driven segmentation system has been developed in the System and Research Center's Image Research Laboratory. The system has shown good results on a number of images from a wide range of scenarios. An overview description of the implemented system is presented in the remainder of this section. Figure 1 shows a block diagram of the prototype knowledge based segmentation system. The processing is divided among four modules which do: goal determination, preprocessing, segmentation, and region evaluation. Each of these modules has a separate knowledge base that contains the production rules specific to the module, and has access to the GKB for scene information.

The goal determination module determines the segmentation goals from the system goals. The system goals are usually based on the specific mission, but they could also conceivably change during a mission. An example goal determination rule from our prototype system is:

RULE #15: IF (GOAL=TARGET_RECOGNITION) THEN (SHAPE_GOAL=CONVEX)

This rule suggests that if the system goal is to recognize targets then the shapes sought will be convex.

The preprocessing algorithm shown in Figure 2, takes in the original image and preprocesses it to accomplish tasks such as noise cleaning and image enhancement. This module uses knowledge, such as sensor type, to determine the preprocessing steps. The preprocessing module can also generate knowledge, such as the overall image contrast, to be used by other modules. A sample preprocessing rule from our prototype system is:

RULE #08: IF (HIGH_FREQ_NOISE) THEN RUN(WINDOW_AVERAGE_ROUTINE)

For this rule it is not important how it was determined that the image has high frequency noise present. It is just important that the noise is present. The information could have come from image measurements or from prior reasoning based on the sensor type and the time of day. After preprocessing comes the actual segmentaton.

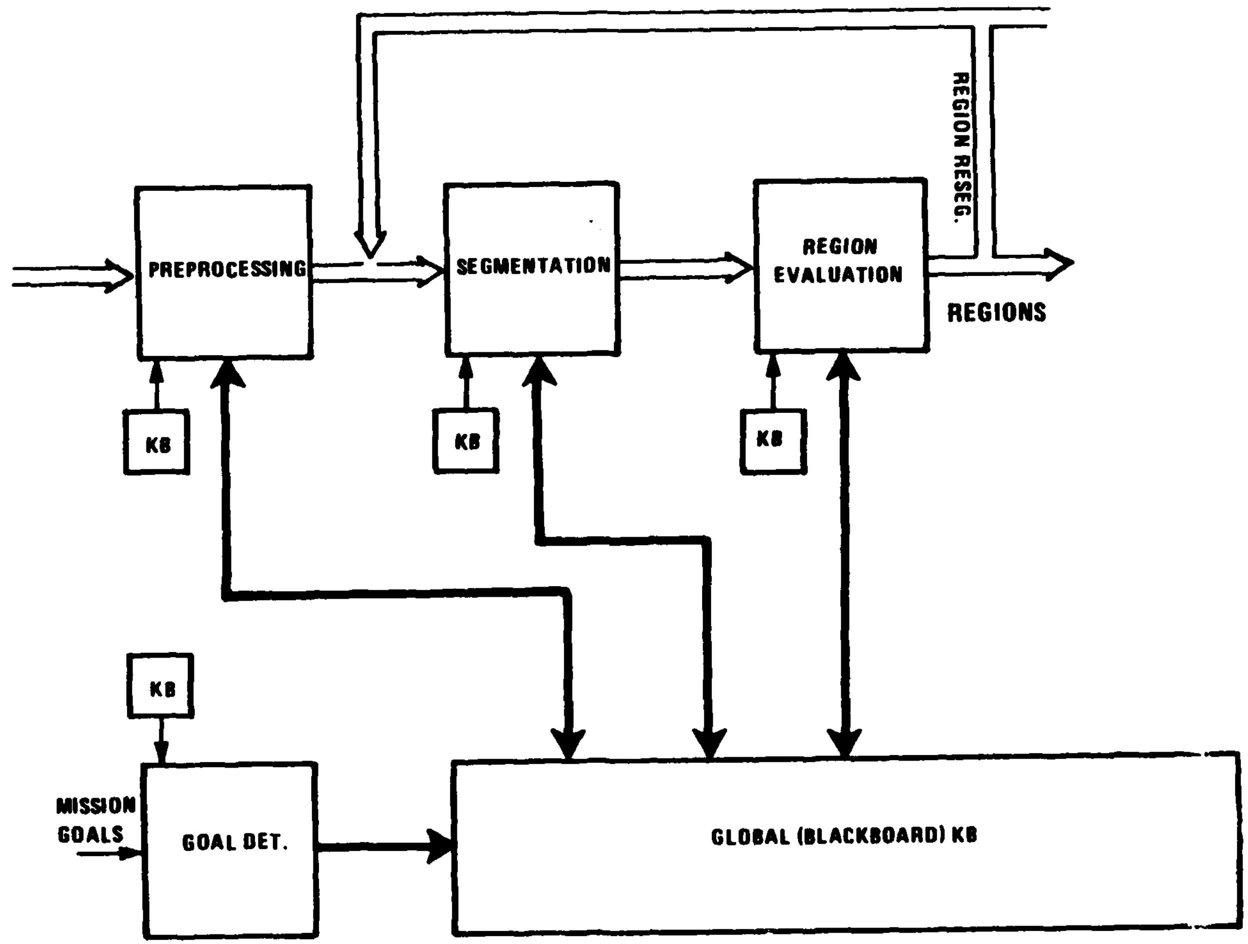


Figure 1. Knowledge based segmentation system

The segmentation module, shown in Figure 3, takes in the preprocessed image and applies segmentation operators from a library of possible operators. Each operator in the library has strong points and weak points. Table 1 shows the suite of segmentation operators which are currently in the system. It is the job of the segmentation module to choose the best operator based on the segmentation goals and on the current scene knowledge from the GKB. Besides doing the actual segmentation, the segmentation module generates knowledge to be put in the GKB for use by other modules. This includes image knowledge, such as the total number of regions segmented, as well as knowledge about each individual region, such as shape and contrast. The segmentation module is sufficiently general so that it can segment regions or entire scenes. This makes the same module usable for initial segmentation of full images or resegmentation of individual regions. An example segmentation rule from a prototype system is:

RULE #47: IF (SHAPE_GOAL=CONVEX and CONTRAST<LOW) THEN (RUN TBL)

This rule recommends the segmentation operator TBL when the contrast of the image is low and the goal is to find convex objects. After the image is segmented into regions, the regions are passed one at a time to the region evaluation module.

SEGMENTOR	TECHNIQUE	APPLICATION
TEXTURE BOUNDARY LOCATOR (TBL)	TEXTURE BASED	LOW CONTRAST BOUNDARIES
PROTOTYPE SIMILARITY TRANSFORM (PST)	REGION BASED	LOW TEXTURE REGIONS
M70 DIGITAL GRADIENT (MDG)	EDGE BASED	HIGH CONTRAST BOUNDARIES

TABLE 1. SEGMENTATION TECHNIQUES USED BY THE KNOWLEDGE BASED SEGMENTATION SYSTEM.

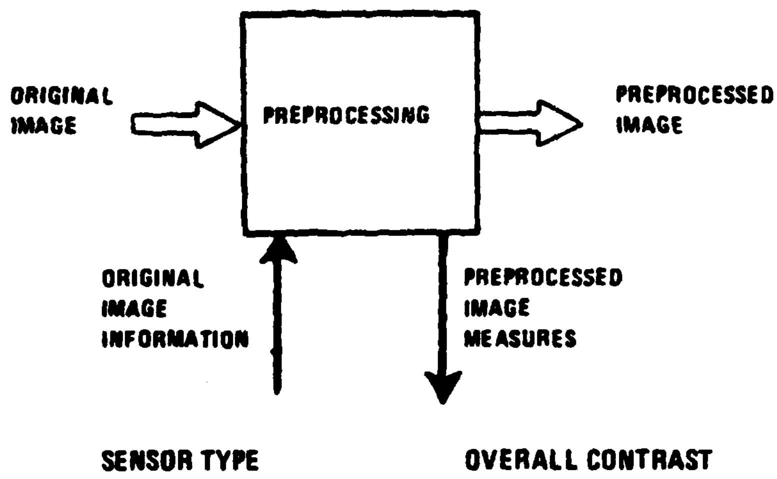


Figure 2. Preprocessing module

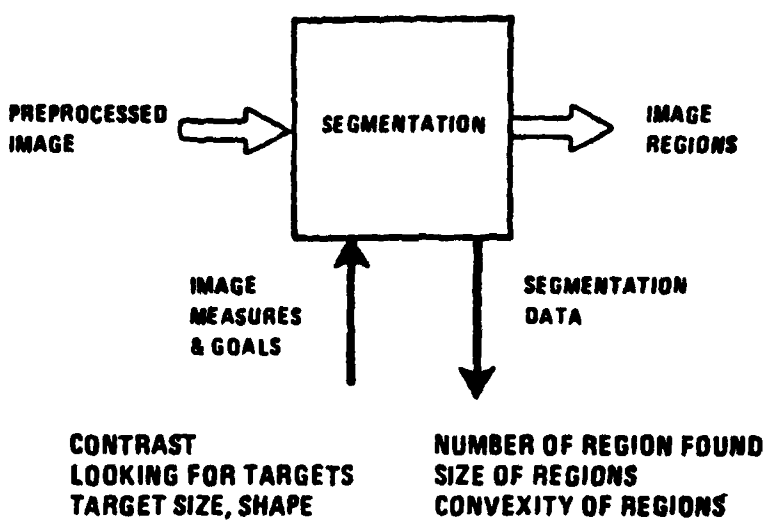


Figure 3. Segmentation module

The region evaluation module, shown in Figure 4, determines what processing should happen next for each region found by the segmentation module. The region evaluation module recommends that a region be: accepted, resegmented, or ignored. This recommendation is based on knowledge about the specific region and on the segmentation goals. The segmentation module, working together with the region evaluation module, produces the best possible segmentation for the image based on the current segmentation goals. An example region evaluation rule is:

```
RULE #75:  IF (GOAL_SIZE=SMALL and GOAL_SHAPE=CONVEX
           and REGION_SIZE=LARGE and REGION_SHAPE=CONCAVE
           THEN (RESEGMENT_REGION)
```

This knowledge driven architecture makes the best use of all available knowledge about the scene. It also provides an approach for making use of information across multiple sensors. The processing for each of the sensors can be designed to be knowledge driven, with all sensor processing paths sharing the same GKB. This would allow each sensor's processing flow to use any knowledge possible, even if it comes from another sensor.

This type of system architecture also can be designed to make use of information across multiple frames of a scene. Information learned from the processing of one frame could be used to help processing of successive frames. For example, if a convex object is found in one frame at a certain size, this infomation could be used to look for more convex objects of the same size in successive frames. This could improve overall system performance by giving the system more information about what it can expect to see in a given frame.

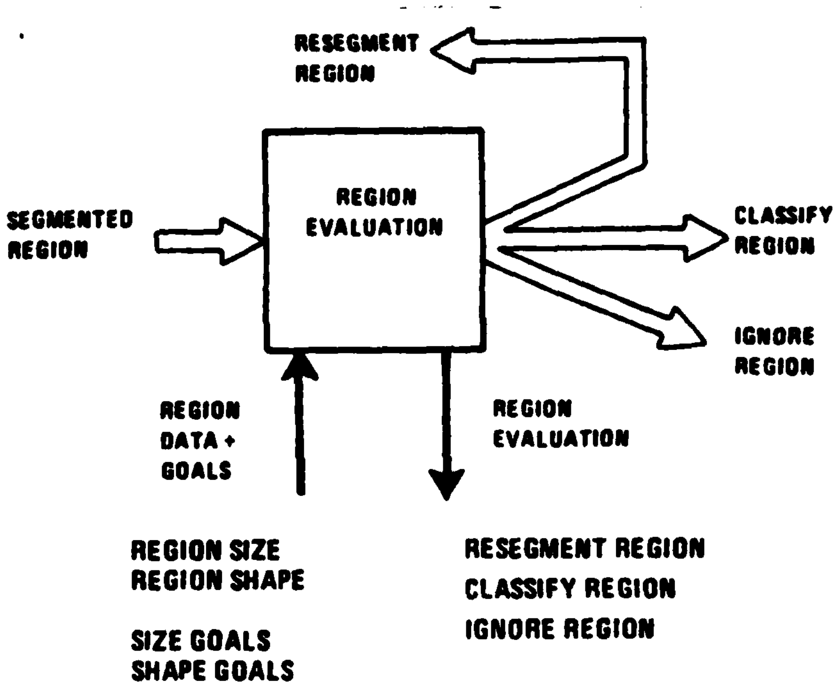


Figure 4. Region evaluation module

Figure 5. Original visible image for a road following scenario

4.0 EXPERIMENTS AND RESULTS

In the following we describe the results of the implemented knowledge based segmentation system for a sample experiment conducted in the Image Research Laboratory for ground scenarios imagery.

Figure 5 shows a sample video image of a road with obstacles on it. This image represents a road following scenario where the goal is to find the road and then determine if there are any obstacles in the road. Figure 6 shows the result of a texture based operator on the road image. Figure 7 shows the results of an edge based segmentor and Figure 8 shows the results of a region based segmentor. None of these results are sufficient to identify the road and the obstacles. The knowledge based control breaks the path following scenario into two goals. The first goal is to find the road by looking for large regions without looking for details. Using these goals the segmentation module recommends running TBL on a lower resolution of the image. Figure 9 shows the results of this first pass of segmentation. Each of the larger regions pass the segmentation goals at this point and are accepted by the region evaluation module.

The second set of segmentation goals for this path following scenario recommends to look for small convex obstacle like regions within each of the larger regions. When the regions are evaluated with respect to these new goals, the region evaluation module recommends passing the larger regions back to the segmentation module for resegmentation. Figure 10 shows the resegmentation results for the center road region. Notice the obstacles are now segmented out. Figure 11 shows the total image segmentation when the resegmentation results are combined backinto a full image.

Figure 6. TBL segmentation of full image

Figure 7. MDG segmentation of full image

Figure 8. PST segmentation of full image

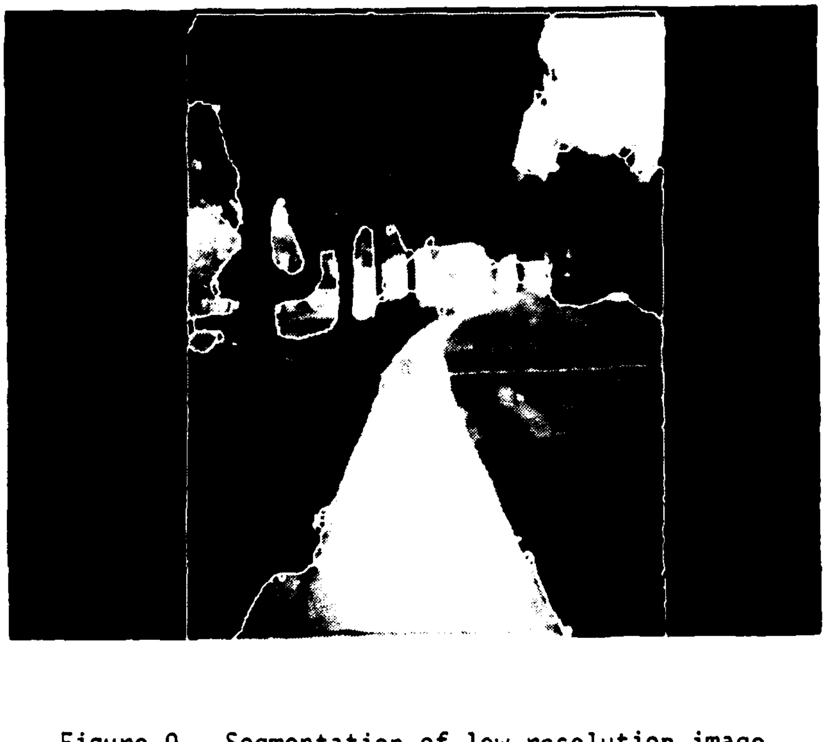


Figure 9. Segmentation of low resolution image

Figure 10. Resegmentation of road region

Figure 11. Full knowledge based segmentation results

5.0 SUMMARY

The overall performance, robustness and multi-scenario operaton of a computer vision system can be improved by making use of all possible knowledge about a scene. Knowledge can be better used by designing the processing control to make use of all available information. By using production rules the system can be knowledge driven and also be flexible and easily expandable when new image processing routines are added to the system. A prototype experimental system has been implemented for knowledge based region segmentation and resegmentation with encouraging results.

A much more difficult problem, than region segmentation, is the classification and labeling of the regions in a scene. The approach we have described in this paper is goal driven segmentation and therefore provides labeling information about the region if the goal is met satisfactorily. This aspect of the approach is being explored as a follow-up to the current application.

REFERENCES

[1] M.E. Bazakos and D.P. Panda, "Stereopsis and Scene Partitioning for Terrain Interpretation," Proceedings of SPIE Conference on Applications of Artificial Intelligence II, Arlington, Virginia, April 1985.

[2] G.S. Duane, S.F. Venable, D.J. Richter, and A.M. Wiedermann, "A Production System for Scene Analysis and Semantically Guided Segmentation", Proceedings of SPIE Conference on Applications of Artificial Intelligence II, Arlington, Virginia, April 1985.

[3] A.M. Nazif and M.D. Levine, "Low Level Image Segmentation: An Expert System", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6 No. 5 Sept. 1984.

THE USE OF OPTICAL FLOW AS A DEPTH CUE IN DYNAMIC SCENE ANALYSIS

TABLE OF CONTENTS

ABSTRACT

This report provides a brief summary of a research project currently under way in the Honeywell Systems and Research Center on the use of "depth cues" in general and on optical flow in particular for dynamic scene analysis. The 2-D optical flow fields are determined by implementing a gradient technique which relates the changes of brightness in the image sequence to the spatial movements in the scene. The results of the use of this technique on a sequence of visible images are given. Furthermore, a multi-sensor approach (which is currently being implemented) is presented for the construction of 3-D optical field from 2-D optical field and range information available from a ranging sensor. This will permit derivation of the complete 3-D motion parameters of the moving objects arbitrary motions in the scene.

1.0 INTRODUCTION

This report summarizes the preliminary results of Internal Research and Development (IR&D) work on the feasibility of the use of 'depth cues' in dynamic scene analysis.

Humans readily use monocular depth cues, such as occlusion, texture gradient and optical flow. [1, 2] Isolated attempts at using some of these depth cues for computer vision and scene understanding have been made by various researchers [3-19] in the last decade. However, no systematic effort on studying the use of these depth cues for computer vision has been reported.

Some of the depth cues such as occlusion and texture gradient are useful in static scenes. Others, such as optical flow, are beneficial in dynamic environments where the position of the viewer with respect to a static scene changes, or the viewer is stationary and the positions of some of the objects in the scene vary with respect to the viewer, or there is a combination of these two cases.

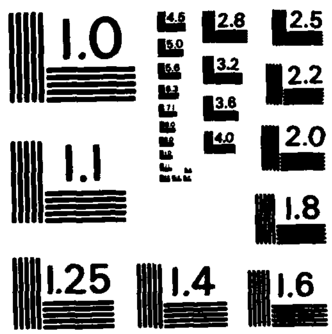
In our study we are considering the use of occlusion cues, texture gradient, and optical flow for the extraction of three-dimensional (3D) information from a scene. In this preliminary report we describe our approach for using optical flow to extract 3D information. This will greatly aid the process of dynamic scene analysis, and will facilitate studies of temporally changing object features. In later reports, we will present the results of our studies in using both occlusion and texture gradient cues.

There are several ways in which the Hierarchical Multisensor Image Understandig (HMIU) program can benefit from these studies. Depth cues in general and optical flow is particular provides 3D information at the lower level of the hierarchy which will be crucial in single sensor image understanding. Moreover, in our approach we plan to initiate a multisensor fusion process at this low level by combining 2D optical flow (obtained from IR or TV) with the range information available from a ranging sensor to create 3D optical flow of the scene. This 3D optical flow is then used to extract 3D motion parameters of the moving objects. We have to note that by using single sensors, only the 2D projected motion parameters can be derived. However, in our multisensor approach, the complete 9 parameters needed for 3D motion description of moving objects can be extracted.

Figure 1 shows the various steps involved in our approach. From the infrared or visual images, 2D optical flow will be derived which will then combine with the range information available from the millimeter wave radar (or lidar) and will lead to the formation of the three-dimensional (3D) optical flow. The nine motion parameters are obtained from each velocity vector. By using the generalized Hough transform method the motion parameters of the rigid body motions in the scene can be specified. In this method, each velocity vector 'votes' for a set of motion parameters and the parameter values receiving the most votes are selected to describe motion of the moving object.

AD-A160 324   HIERARCHICAL MULTISENSOR IMAGE UNDERSTANDING(U)   2/2
HONEYWELL SYSTEMS AND RESEARCH CENTER MINNEAPOLIS MN
R K AGGARWAL ET AL. AUG 85 AFOSR-TR-85-0801
UNCLASSIFIED   F49620-83-C-0134   F/G 12/1   NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

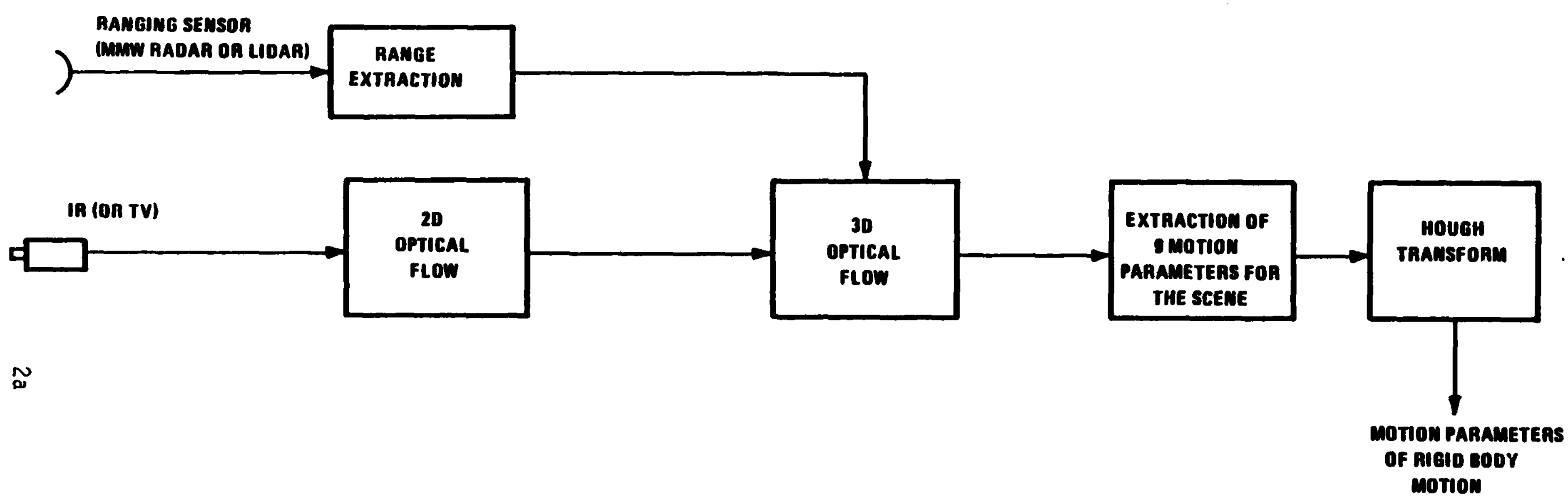


Figure 1. Honeywell's Approach for the Extraction of Moving Objects' Motion Parameters.

2.0 DETERMINING 2-D OPTICAL FLOW

There are two main approaches in the literature for the determination of optical flow: gradient methods and matching techniques. In matching techniques, one has to follow the motion of prominent feature points from frame to frame in an image sequence. These techniques require an initial segmentation of the imagery and more importantly they rely on the finding of the best matching points. The gradient methods, however, avoid the correspondence problem by using the original unsegmented images and making use of the gradient constraint equation which relates the changes in brightness to the temporal changes in the scene. For an object of constant brightness U(x, y, z), the following equation is derived

$$V \cdot \nabla u + \frac{\partial u}{\partial t} = 0 \tag{1}$$

where

$$V \equiv \left[\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}\right]^t = \left[v_x, v_y\right]^t \tag{2}$$

$$\nabla u \equiv \left[\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right] \equiv \left[u_x, u_y\right] \tag{3}$$

We used a gradient method [7,13] for the determination of optical flow. This technique is an iterative and robust method for the determination of 2D optical flow V. Some of the features of this method are as follows:

- Samples are taken at discrete points in space and time and quantized in brightness.
- Partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ and $\frac{\partial u}{\partial t}$ are estimates by averages using eight measurements in two image frames.

o The criteria that is minimized is the square of the magnitude of the gradient of the optical flow velocity components v_x, v_y :

$$\left(\frac{\partial v_x}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial y}\right)^2 + \left(\frac{\partial v_y}{\partial x}\right)^2 + \left(\frac{\partial v_y}{\partial y}\right)^2 \qquad (4)$$

The final iterative relations for estimating v_x and v_y are as the following:

$$v_x^{n+1} = \bar{v}_x^{\,n} - u_x\left[u_x \bar{v}_x^{\,n} + u_y \bar{v}_y^{\,n} + u_t\right] / (\alpha^2 + u_x^2 + u_y^2) \qquad (5.a)$$

$$v_y^{n+1} = \bar{v}_y^{\,n} - u_y\left[u_x \bar{v}_x^{\,n} + u_y \bar{v}_y^{\,n} + u_t\right] / (\alpha^2 + u_x^2 + u_y^2) \qquad (5.b)$$

where α is a weighting factor.

3.0 DETERMINING 3D OPTICAL FLOW

Consider a viewing geometry, as depicted in the Figure 2, defining the three components of 3D optical flow as W_x, W_y and W_z, and, using geometrical relations, the following equation can be derived:

$$(f-z)V_x = W_x f + x^1 W_z \quad (6\cdot a)$$

$$(f-z)V_y = W_y f + y^1 W_z \quad (6\cdot b)$$

By using focal length of the camera f; positional coordinates in the image x^1, y^1; 2D optical flow velocity components V_x and V_y; and depth information z, W_z; the components W_x and W_y will be computed.

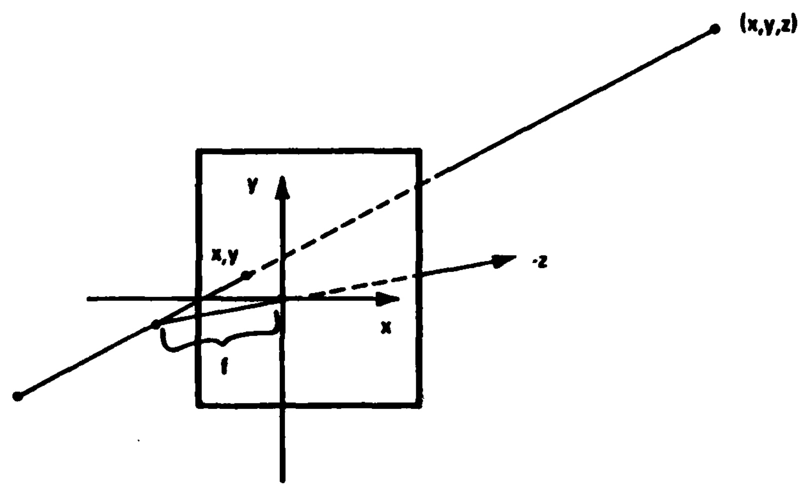


Figure 2. The Viewing Geometry for Determining 3D Optical Flow from 2D Optical Flow.

4.0 DETERMINING MOTION PARAMETERS

The motion of rigid bodies in the scene are described relative to a frame of reference. We assume that this frame is fixed to the viewer. The most general type of motion is described by the following relation:

$$X = X_A + P \qquad (7\cdot a)$$

$$V = V_A + \quad XP \qquad (7\cdot b)$$

where $P = X - X_A$

and Ω is the rotation vector at a point A on the rigid body. V_A is the velocity component associated with the translation velocity of the point A. so the rigid body motion can be described by 9 parameters V_A, Ω, X, X_A. Figure 3 shows the relations between the viewer frame of reference and the rigid body's frame of reference. For determining Ω we first calculate rotational direction W which is defined as $\frac{\Omega}{|\Omega|}$ by using three successive 3D optical flow measurements to obtain 2 acceleration vectors a_1, a_2 W is found from the following relations

$$W = a_1 \times a_2 / (a_1 \times a_2) \qquad (8)$$

V_T and $|\Omega|$ will be determined from the following relation:

$$V = V_T + |\Omega| W \times P \qquad (9)$$

From each velocity vector the motion parameters will be extracted and then by using a Hough transform [20] like procedure, the motion parameters of the solid objects will be extracted by noticing that moving rigid bodies will have the highest number of points having the same motion parameters in the scene. The Hough transform was originally developed for the detection of straight line segments in image. In this technique the points in the image

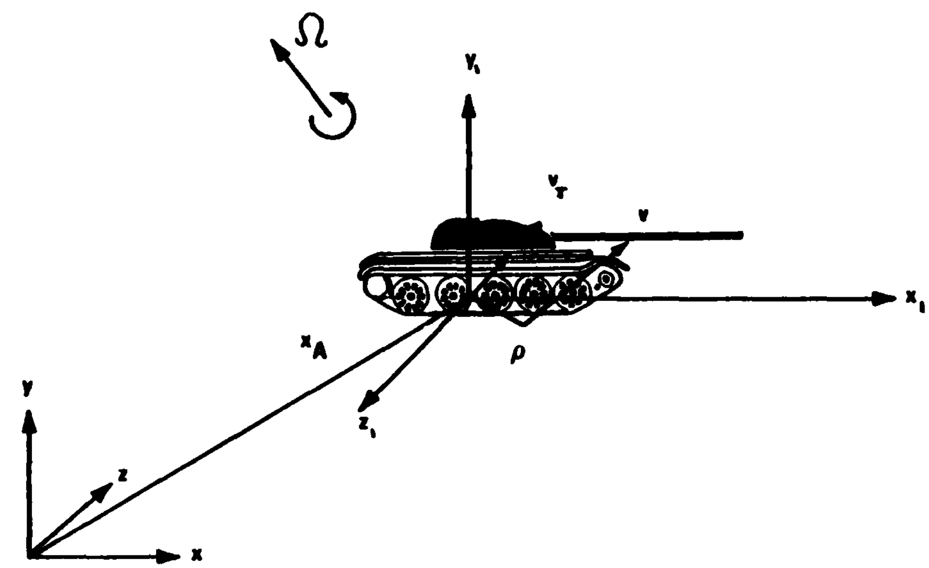


Figure 3. Relations Between Viewer and Rigid Body Reference Frames.

are transformed into lines in a parametric (line-intercept) space. Lines in the slope-intercept space corresponding to colinear points will cross each other in one point. This technique recently has been generalized for detection of more complex patterns in multi-dimensional spaces.

5.0 PRELIMINARY EXPERIMENTAL RESULTS

A sequence of 12 TV images were created from a moving toy jeep at the rate of 30 frames per second. Figures 4 and 5 show the first two frames of pictures of the moving jeep. Using the gradient technique described previously, the optical flow fields by using different number of frames and varying the values of α, N and the scale factor were obtained. Figure 6 shows the derived optical flow fields for $\alpha = 2.0$, N = 16 and scale factor 2. As can be seen from Figure 5, the outer boundaries of the jeep can be easily inferred from the optical flow fields. Furthermore, some of the surface orientation properties of the jeep have also exhibited themselves in the optical flow field. This is very useful information that is currently under investigation.

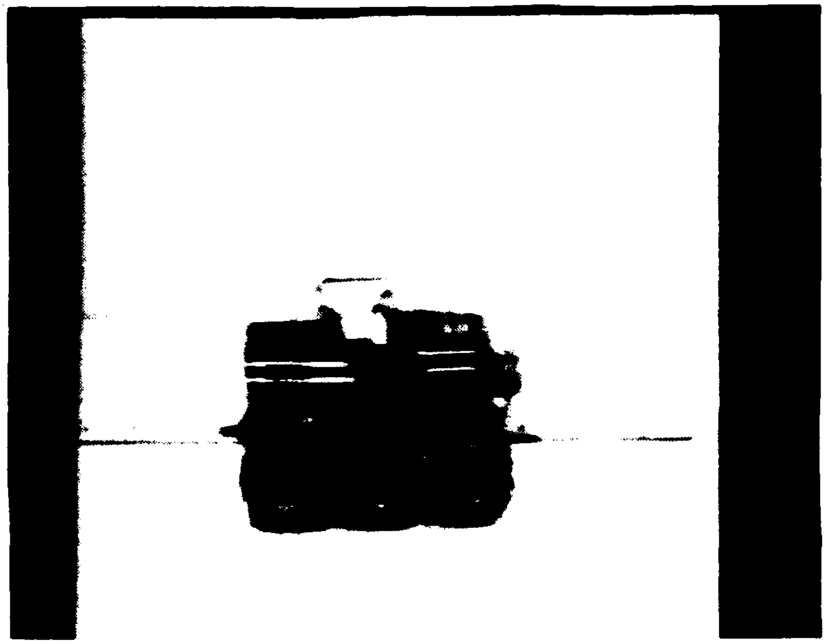


Figure 4. Image of the First Frame of a Sequence of a Moving Jeep

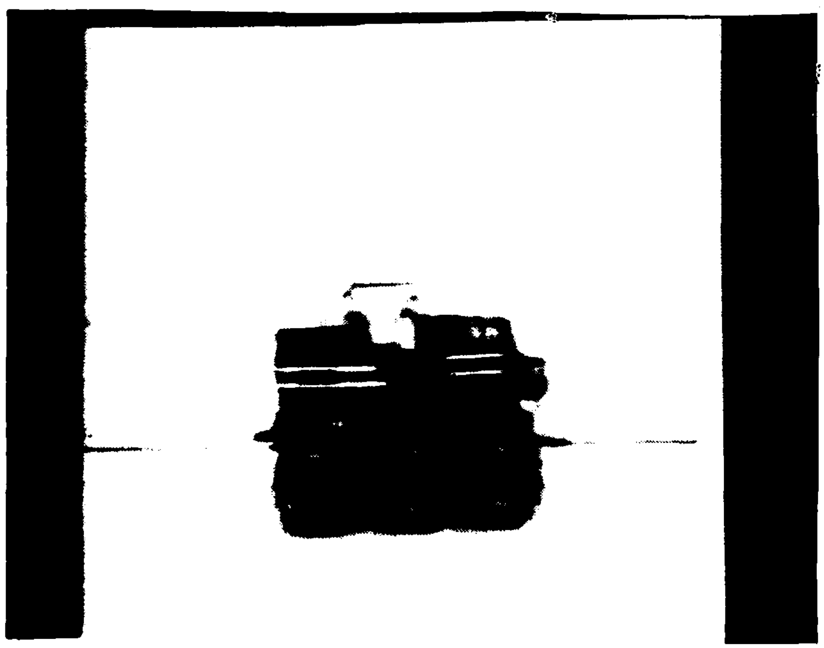


Figure 5. Image of the Second Frame of a Sequence of a Moving Jeep

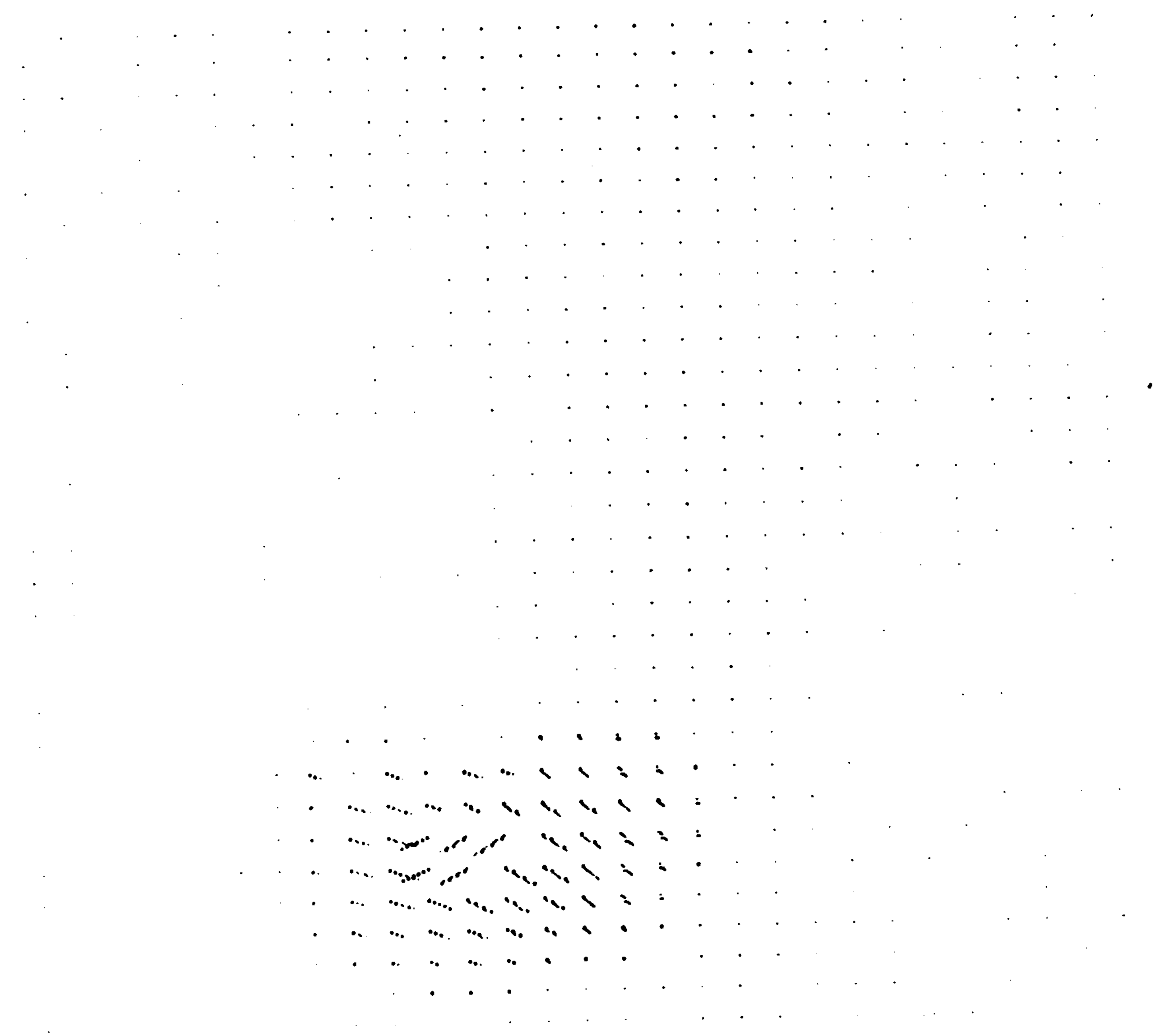


Figure 6. Optical Flow Field of Jeep Using the First Two Frames of Pictures

6.0 SUMMARY

In this report a brief description of a research project currently under way at the Honeywell SRC/SIP on the use of "depth cues" in general and optical flow in particular for dynamic scene analysis was given. Our overall multi-sensor approach for the construction of 3-D optical flow and the extraction of 3-D motion parameters was described and the preliminary results of the implementation of a 2-D optical flow technique as part of our overall approach, on a sequence of visible images were presented.

References

[1] Gregory, R.L., "Seeing in Depth", Proceeding of SPIE: Three-Dimensional Imaging, Vol. 120, 1977.

[2] Nakayama, K., "Geometric & Physiological Aspects of Depth Perception," Proceedings of SPIE: Three-Dimensional Imaging, Vol. 120, 1977.

[3] Williams, T.D., "Computer Interpretation of a Dynamic Image from a Moving Vehicle", Doctoral Dissertation, Massachusetts University, Amherst, Mass., 1981.

[4] Boes, U., "Detection of Moving Objects by Space and Time Invariant Filters," Report No: ESA-TT-753, European Space Agency, Paris, France, 1983.

[5] Prazdny, K., "Egomotion and Relative Depth Map from Optical Flow,", Biol. Cybern (Germany) Vol. 36, No. 2, 1980.

[6] Prazdny, K., "Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinearly Moving Observor," Pro. SPIE: Tech. & Appl. Image Understanding, Vol. 281, 1981.

[7] Paquir, R. Dubois, E. "A Spatio-Temporal Gradient Method for Estimating the Displacement Field in Time-Varying Imagery," Computer Vision, Graphics & Image Processing, Vol. 21, No. 2, 1983.

[8] Neumann, B. "Optical Flow," Computer Graphics, Vo. 18, No.1, 1984.

[9] Buxton, B.F., Buxton, H., "Low-level Edge Detection for the Extraction of Optical Flow Data and Passive Ranging," GEC Journal of Research, Vol. 1, No. 3, 1983.

[10] Wiejak, J.S., "The Correspondence Problem in Stereopsis," IEE Colloquium on Scene Matching in Multi-site Sensors, 1984.

[11] Rosenberg, D., Levine, M.D., " Computing Relative Depth from Occlusion Cues," Proc. 4th International Joint Conference on Pattern Recognition, 1978.

[12] Bajcy, R., "Texture Gradient as Perception Cue", Comput. Graph Image Process, 1976.

[13] Horn, B.K.P., Schunck, B.G., "Determining Optical Flow," Proc. SPIE: Tech. & Appl. Image Understanding, Vol. 281, 1981.

[14] Prazdny, K., "On the Information in Optical Flows," Comput. vision, Graph., Image Process, Vol. 22, 1983.

[15] Ballard, D., Kimball, O., "Rigid Body Motion from Depth and Optical Flow," Computer Vision, Graphics and Image Processing, Vol. 22, 1983.

[16] Bazakos, M., D. Panda, D. Duncan, "Passive Ranging in Outdoor Environment," Proceedings of IEEE Conf. on Robotics & Automation, April 1985.

[17] Barnard, S.T., Thompson, W.B., "Disparity Analysis of Images," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-2, pp. 333-340, July 1980.

[18] Kearney, J.K., Thompson, W.B., "Gradient-based Estimation of Optical Flow with Global Optimization," Proceedings of the 1984 IEEE Conf. on Artificial Intelligence Applications, Denver, CO, December 1984.

[19] Murray, D.W., Buston, B.F., "Reconstructing the Optic Flow Field from Edge Motion: An Examination of Two Different Approaches," Proceedings of 1984 IEEE Conf. on Artificial Intelligence Applications, Denver, CO, December 1984.

[20] Hough, P.V.C., Method and Means for Recognizing Complex Patterns, U.S. Patent 3069654, 1962.

TR85.5

BELIEF MAINTENANCE FOR A FUZZY REASONING SYSTEM

BELIEF MAINTENANCE FOR A FUZZY REASONING SYSTEM

ABSTRACT

Real world problem solving often involves (1) dealing with uncertain and imprecise knowledge and (2) making assumptions which are then verified or denied by the reasoning process. Fuzzy logic is presented as the mechanism for dealing with measures of belief and a maintenance system is proposed for handling assumptions and the accrual of evidence.

TABLE OF CONTENTS

BELIEF MAINTENANCE FOR A FUZZY REASONING SYSTEM

I.0 INTRODUCTION

Expert systems and the accompanying reasoning systems are at the forefront of the artificial intelligence research that is being applied to computer vision and image understanding systems [1,5,7,9,11,16]. The difficulties associated with acquiring and encoding expertise are a major bottleneck in the practical application of expert systems. Expert knowledge is often imprecise (vague) or uncertain and the rules used in inferencing with knowledge in this form must model and support measures of belief. Several methods based on probability have been presented and used for dealing with uncertainty [2,6,15], yet none of these methods deals well with the imprecision or vagueness inherent in expert knowledge. Fuzzy logic, based on fuzzy set theory [19,21,22], assigns intervals of possible values to a fact rather than a single probabilistic value. Research is being pursued in unifying the possibilistic and probabilistic approaches [8,13,18] and in the application of fuzzy logic to expert systems [1,8,12,17,20] and to image understanding [14]. Section II explains the basic concepts of inferencing with fuzzy logic.

The knowledge in reasoning systems is also incomplete and assumptions must be advanced or it is inconsistent and conflicts must be resolved. Human reasoning makes use of the assumption verification/denial paradigm. In problem solving, the human will assume certain facts in order to drive the reasoning process. If an assumption is proven incorrect, the reasoning based on that assumption is ignored and the problem solving continues, possibly with another assumption. This sort of non-monotonic reasoning requires facilities for maintaining evidence for facts and resolving conflicts [10]. In [3], facts are based on endorsements and in [4] facts are considered true based on a support list. The reader should investigate these methods; however, an extension to the latter method is provided in Section III.

Section IV is an example of a belief maintenance system for fuzzy logic applied to the classification of regions within a scene. The system maintains dynamically accrued facts and assumptions with measures of belief.

2.0 FUZZY SET THEORY AND FUZZY LOGIC

Fuzzy set theory, introduced by Zadeh [19] in 1965, deals with the notion of imprecision in discrimination between sets of objects. Conventional or 'crisp' sets have sharp boundaries and objects either belong to a specific set or do not belong to that set. Fuzzy sets, on the other hand, contain objects which have a degree of membership between 1 (full membership) and 0 (nonmembership). An example may further clarify the difference. The crisp set of BRIGHT pixels may be defined as all pixels where the average pixel intensity is above some threshold. The fuzzy set of BRIGHT pixels assigns a degree of membership to each pixel value, such that pixels with a low intensity have a low degree of membership in the set of BRIGHT pixels and vice versa for pixels of high intensity.

Let S be a set of objects and F be a fuzzy subset of S, such that for each s in S, there is an associated degree of membership in F. If S is the set of pixel intensities (values 0 to 255) then BRIGHT could be a fuzzy subset of S, where each pixel intensity in S has a degree of membership in the set BRIGHT. In the following example of a four interval set BRIGHT, the degree of membership is .3 for a pixel intensity of 120.

$$\text{BRIGHT} = 0/\{0\text{-}63\} + .3/\{64\text{-}127\} + .7/\{128\text{-}191\} + 1.0/\{192\text{-}255\}$$

If the fuzzy subset is continuous rather than discretized into intervals, the membership function becomes a curve as in Figure 1.

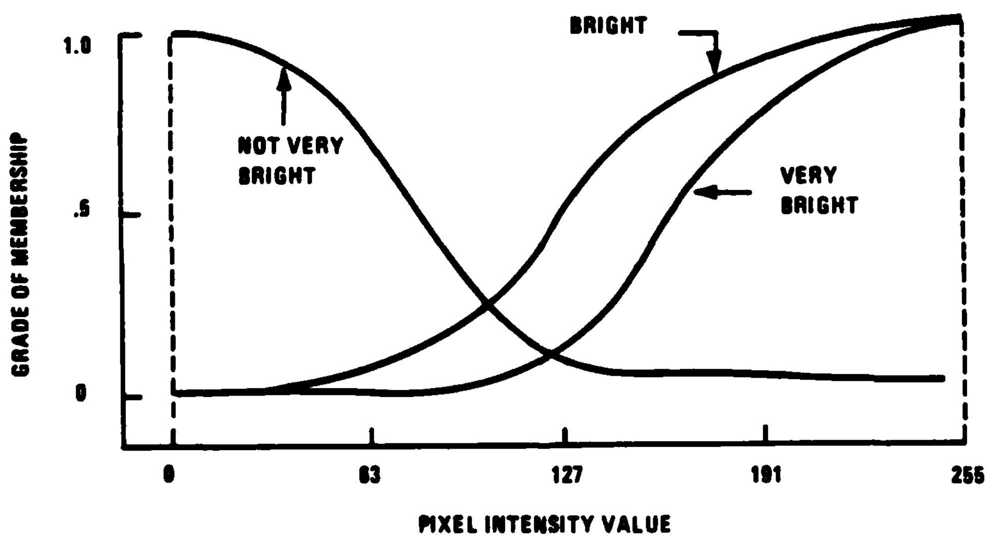


Figure 1

In a rule based system the predicates or facts as well as the antecedents and consequents of the rules are in the form:

"attribute of object IS/HAS value".

If V is a variable for the attribute of an object where the domain of V is S, then the proposition "V is F", where F is a fuzzy set, induces a possibility distribution, u, over the set S, such that u(s) = F(s). F(s) is the degree of membership of s in F and u(s) is the possibility that V = s given the data "V is F". The differences between possibility values and probability values is heavily debated and not an issue here, yet an example may provide some insight into the difference. In the above fuzzy set BRIGHT we see that pixels with an intensity value between 128 and 191 are given a .7 degree of membership in BRIGHT. This can be viewed as a 70% possibility that a pixel intensity is BRIGHT, yet it cannot be interpreted as .7 probability. In probability, the sum of all the individual probabilities is 1, while in possibility many set members may have the value of 1 (full membership). Therefore the proposition "V is F" assigns a possibility or degree of membership based on the fuzzy set F, to each value V can represent.

Rules are represented in a conditional form such as

"if V is F then W is G"

where F and G are fuzzy subsets of S and T respectively. This rule induces a conditional possibility distribution u(s,t) where

u(s,t) = min (1, 1 - F(s) + G(t)).

Compound antecedents or consequents are dealt with by two other compositional rules. The conjunction form

"V1 is F1 and V2 is F2 ... Vn is Fn"

where Fj is a fuzzy subset of Sj, induces a joint possibility distribution u(s1,s2,...sn) where

u(s1,s2,...sn) = min [Fj(sj)] for j=1,n.

The disjunction form

"V1 is F1 or V2 is F2 ... Vn is Fn"

where Fj is a fuzzy subset of Sj, induces a joint possibility distribution u(s1,s2,...sn) where

$$u(s1,s2,...sn) = \max [Fj(sj)] \text{ for } j=1,n.$$

Combinations of these compositional rules and the conditional rule can be used to generate possibility distributions. For example,

"if V1 is F1 or V2 is F2 then W is G"

induces the possibility distribution u(s1,s2,t) where

$$u(s1,s2,t) - \min (1, 1 - \max [F1(s1),F2(s2)] + G(t)).$$

There are also modification rules such as "not" and "very".

"V is not F" induces $u'(s) = 1 - F(s)$.
"V is very F" may induce $u''(s) = [F(s)]$ squared.

Inferencing about fuzzy knowledge is based on modus ponens. That is, given A and the knowledge that A implies B, then B is inferred. This is a conjunctive of the form

"V is F and (if V is F then W is G)"

where F and G are fuzzy subsets of S and T respectively. This induces the joint possibility distribution u(s,t) where

$$\begin{aligned} u(s,t) &= \min [F(s), \min [1,1-F(s)+G(t)]] \\ &= \min [F(s), 1-F(s)+G(t)]. \end{aligned}$$

Since the possibility distribution u(t) for "W is G" is what is being inferred, then the maximum u(s,t) for each t is the possibility distribution for each G(t). This leads to

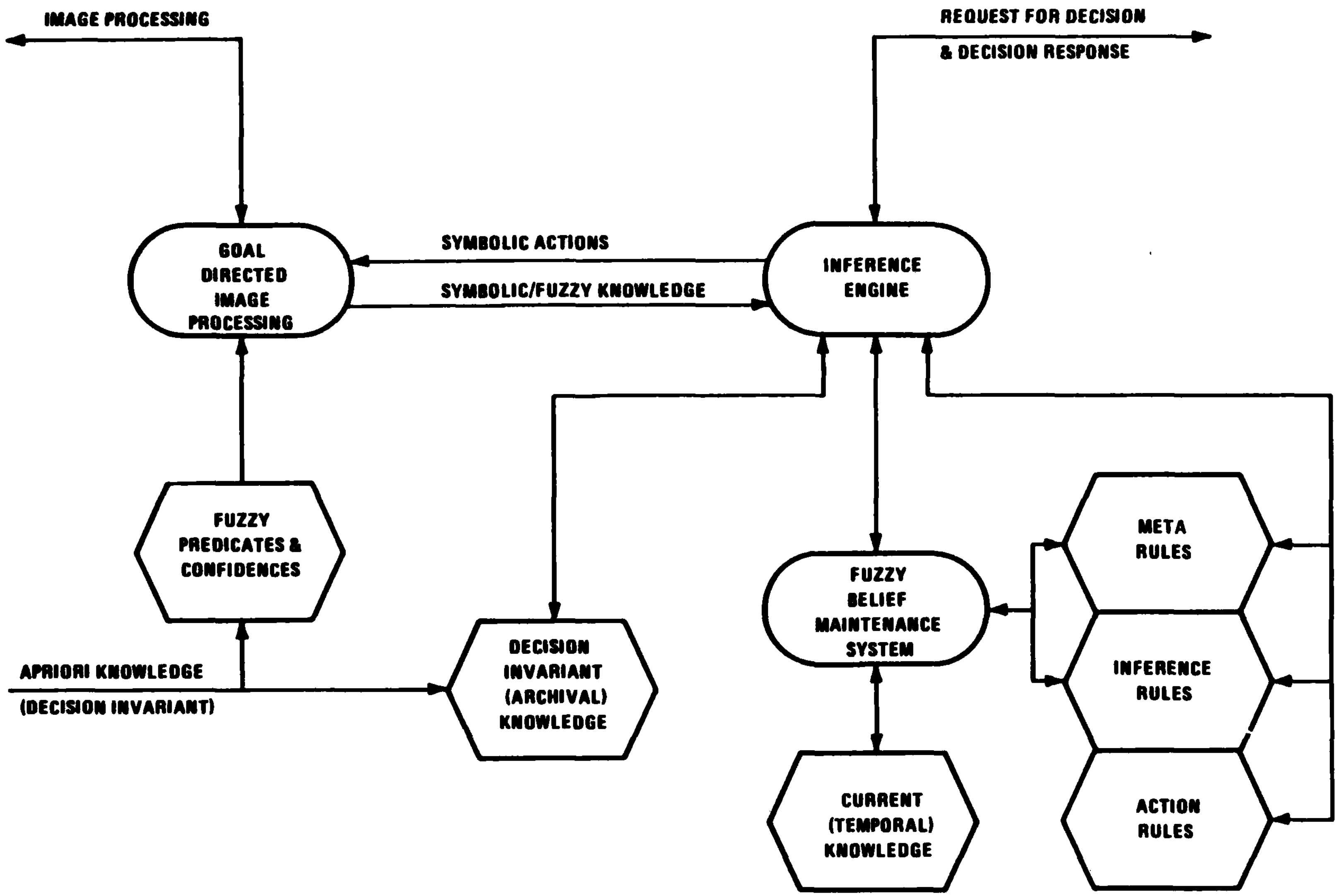
u(t) = maximum u(s,t) over all s
= max [min [F(s), 1-F(s)+G(t)]] over all s.
= min [F(s), 1-F(s)+G(t)] when s is known

The strength of a fuzzy set value can be determined in several ways.

1. The expected value may be obtained by some method such as mean distribution and the strength of a fuzzy set is the membership value at that expected value.

2. The strength could also be the sum of the membership values for each interval. This is the integral of a continuous function over the fuzzy set.

3. A threshold membership value is determined and some measure of those intervals which exceed the threshold is used as the strength of the fuzzy set.



PROCESS FLOW DIAGRAM FOR A
FUZZY BELIEF MAINTENANCE SYSTEM
Figure 2

3.0 BELIEF MAINTENANCE

Figure 2 gives an example of a reasoning system for image understanding. The fuzzy belief maintenance system receives predicates from the inference engine and maintains consistency among beliefs. Section IV gives an example of this.

In reasoning over a set of rules and facts, there may be equally certain yet conflicting facts, or there may not be enough current knowledge to confirm a fact. An assumption is advanced and the reasoning process continues (as may the knowledge acquisition process). At some point in this process, the assumption may be confirmed and all knowledge inferred from the assumption is updated accordingly, or the assumption is denied and the knowledge which is singularly inferred from the assumption is removed from the set of current beliefs.

Inconsistency in the current set of beliefs occurs when a belief or predicate has two conflicting values. Conflict resolution in this sense then encompasses dealing with two confirming pieces of evidence as well as complementary pieces of evidence.

Conflict resolution occurs at the instant a conflict arises. There are four meta-rules for the resolution process.

1. If neither the current belief of a fact nor the new belief of that fact are based on any assumption, then if the new belief is stronger than the current belief (measures of strength discussed above), replace the current belief with the new belief and propagate the new belief value to all the other facts dependent on the fact.

2. If the current belief of a fact is based on an assumption and the new belief of that fact is not, then replace the current belief with the new belief and propagate that value to all the other facts dependent on the fact.

3. If the current belief of a fact is not based on an assumption and the new belief of that fact is, then do nothing.

4. If the current belief and new belief of a fact are based on assumptions, then if the new belief is "significantly less assumption based", replace the current belief with the new belief and propagate the new belief value to all the other facts dependent on the fact.

In the fourth meta-rule above, the qualifier "signficantly less assumption based" is some measure of inherited assumption which is a distance function in the inference network. Intuitively, beliefs based on some distant assumptions may also have inherited some facts, while beliefs based on less distant assumptions are likely to have inherited less facts. Practically, retaining the most distant belief will mean less propagation.

The format for beliefs being maintained is:

(FACT) (FUZZY SET) (BASIS) (VALUE)

where FACT is a predicate from the knowledge base (ie, REGION_INTENSITY is BRIGHT);

FUZZY SET is the fuzzy set associated with the fact (ie, the fuzzy set BRIGHT);

BASIS is the support for the fact (this includes a tag for indicating inheritance of assumption, all rules which infer the FACT, and the accompanying facts which triggered the rules);

VALUE is the real value of the fact (ie, 128 is the REGION_INTENSITY). If VALUE is empty then the expected value of the fuzzy set can be computed and used as the value.

4.0 EXAMPLE OF BELIEF MAINTENANCE FOR REGION CLASSIFICATION

In the following example, the fuzzy techniques and the belief maintenance system, described above, are integrated into a region classification tool. Sensor input to the system is provided by MMW and IR sensors. In addition, wind velocity in scene, sensor inclination (vertical direction), apriori knowledge (ie, map), and region knowledge from previous scenes or from other regions in the current scene, are inputs to the system.

Example Rule Base:

R1 HIGH(LINE OF SIGHT FROM INCLINATION) ==>> (SKY)
R2 (SKY) ==>> UNLIKELY(ROAD) and UNLIKELY(RIVER)
R3 HIGH(TEXTURE FROM MMW) ==>> (ROAD)
R4 LOW(TEXTURE FROM MMW) and (WINDY) ==>> (RIVER)
R5 HIGH(INTENSITY FROM IR) ==>> (ROAD)
R6 LOW(INTENSITY FROM IR) ==>> LIKELY(RIVER)

The quantifiers/qualifiers preceding the predicates, modify the predicates in specified ways (see [18-22]). HIGH may shift all the degrees of membership to the right and LOW may shift all the degrees to the left as in the following example.

domain:pixel values	0-50	51-101	102-152	153-203	204-255
(average) INTENSITY	.1	.7	1	.6	.1
HIGH INTENSITY	0	0	.1	.7	1
LOW INTENSITY	1	.6	.1	0	0

Thus, an IR intensity value of 100 would be a 0 degree of membership in the fuzzy set HIGH INTENSITY, and .6 degree of membership in LOW INTENSITY. LIKELY increases the degree of membership for each interval, while UNLIKELY decreases the degrees of membership. In the following example, ROAD is a fuzzy set over the domain of depth, such that a road is less likely to be classifiable 10 km from a sensor than a road 2 km from a sensor. The following example increases/decreases ROAD by .2 degrees of possibility.

domain:MMW range(km)	1-3	3-5	5-7	7-9	9-11	11-
(default) ROAD	1	1	.8	.6	.3	.1
LIKELY ROAD	1	1	1	.8	.5	.3
UNLIKELY ROAD	.8	.8	.6	.4	.1	0

The predicates and associated fuzzy sets in the example are as follows.

The SKY domain is the mean horizon height in an image (512 lines) such that a region between 350 and 400 pixels has a .8 degree of possibility (pixels numbered from bottom).

SKY =	0-101	102-204	205-307	308-409	410-511
degree	0	.2	.5	.8	1

The ROAD domain is the depth of region such that a region at 8 km has a .6 degree of possibility.

ROAD =	1-3	3-5	5-7	7-9	9-11	11-
degree	1	1	.8	.6	.3	.1

The RIVER domain is the depth of region similar to ROAD.

RIVER =	1-3	3-5	5-7	7-9	9-11	11-
degree	1	.9	.7	.4	.2	.1

The WINDY domain is the velocity at time of image.

WINDY =	05-5	5-10	10-15	15-20	20-25	25-30	30-
degree	0	.1	.3	.5	.7	.9	1

The LINE OF SIGHT (LOS) domain is the inclination in angular degrees of sensor. The possibility of regions belonging to a certain class may be influenced by the line of sight (see rule 1).

LOS =	-30'	-20'	-10'	0'	10'	20'	30'
degree	.1	.5	.8	1	.8	.5	.1

The TEXTURE domain is based on the mean intensity value from MMW in a granularity measure from 1-8, where 1 is low texture.

TEXTURE =	1	2	3	4	5	6	7	8
degree	0	.3	.6	.9	1	.7	.4	.1

The INTENSITY domain is the mean pixel brightness of IR.

INTENSITY =	0-50	51-101	102-152	153-203	204-25[illegible]
degree	.1	.7	1	.6	.1

Given a value for a fact, the fact and current degree of membership for the value are added to the list of beliefs. If no value is known but the fact is either input or inferred then the fact is added to the list of beliefs and the value will be the mean distribution of the fuzzy set.

In the following example, nothing is known at the start of the classification, but as requests for more information are answered by the DASM knowledge is accrued and a classification is arrived at. The system begins with the initial classification RIVER (possibly from apriori knowledge of the scene).

1. (RIVER) (1 .9 .7 .4 .2 .1) (A) ()

Information is requested by the system to continue processing. The information may come from sensors, apriori knowledge, etc. LINE OF SIGHT at a 10 degree incline is input from DASM.

2. (LOS) (.1 .5 .8 1 .8 .5 .1) () (+10')

Rule 1 is triggered, adding SKY to the belief list.

3. (SKY) (.2 .4 .7 .8 .8) (R1 2) ()

4. (ROAD) (.5 .5 .5 .5 .5 .5) (R2 3) ()

Rule 2 infers ROAD above and infers RIVER below.

(RIVER) (.5 .5 .5 .5 .5 .5) (R2 3) ()

Since the original RIVER was assumption based, it is replaced.

1. (RIVER) (.5 .5 .5 .5 .5 .5) (R2 3) ()

The system requests more information from DASM.

5. (TEXTURE) (0 .3 .6 .9 1 .7 .4 .1) () (6)

6. (WINDY) (0 .1 .3 .5 .7 .9 1) () (25 mph)

Rule 3 is now triggered, inferring ROAD again.

(ROAD) (.6 .6 .6 .6 .6 .5) (R3 5) ()

Since the original ROAD and the new ROAD are both fact based, the fuzzy sets are ORed (max of each interval).

4. (ROAD) (.6 .6 .6 .6 .6 .5) (R2 3, R3 5) ()

Rule 4 infers RIVER with a value below some threshold and the new value for RIVER is ignored. RIVER could have ben ORed with the original value producing the same results.

(RIVER) (.1 .1 .1 .1 .1 .1) (R4 5 6) ()

Suppose a strength of .7 is needed to conclude a fact.
ROAD has a strength of .6 and RIVER has a strength of .5 resulting in a request from DASM for more information in order to obtain a stronger conclusion. Here the intensity value from an IR sensor is input to the system.

7. (INTENSITY) (.1 .7 1 .6 .1) () (45)

Rules 5 and 6 are triggered inferring RIVER and not inferring ROAD at all.

(ROAD) (0 0 0 0 0 0) (R5 7) ()

ROAD is ignored since the value or strength is nil.

(RIVER) (1 1 .9 .6 .4 .3) (R6 7) ()

The new RIVER is ORed with the original RIVER resulting in a strong conclusion for river.

1. (RIVER) (1 1 .9 .6 .5 .5) (R2 3, R6 7) ()

At this point, RIVER has a strength of .75 and the particular region is classified as a river.

5.0 SUMMARY

Fuzzy logic provides a tool for belief measures in reasoning systems with uncertain and imprecise facts. The non-monotonic belief system proposed here reasons over incomplete and inconsistent knowledge. Assumptions assist the acquisition of knowledge and can be removed from the set of currently believed facts when the assumptions conflict with non-assumed facts. This belief system enhances the acquisition and inferencing of knowledge in a fuzzy reasoning system.

REFERENCES

[1] Appelbaum, L., and Ruspini, E., Aries: A Tool for Inference under Conditions of Imprecision and Uncertainty, SPIE Applications of Artificial Intelligence II, 1985, pp. 161-165.

[2] Barnett, J., Computational Methods for a Mathematical Theory of Evidence, Proceedings of International Joint Conference on Artificial Intelligence, 1981.

[3] Cohen, P., and Grinberg, M., A Theory of Heuristic Reasoning About Uncertainty, AI Magazine, Vol. 4, No. 2, 1983, pp. 17-24.

[4] Doyle, J., A Truth Maintenance System, Artificial Intelligence, Vol. 12, November 1979, pp. 231-272.

[5] Duane, G.S., Vanable, S.F., Richter, D.J., and Wiedemann, A.M., A Production System for Scene Analysis and Semantically Guided Segmentation, Proceedings of the SPIE Applications of Artificial Intelligence II, 1985, pp. 35-42.

[6] Duda, R., Hart, P., and Nilsson, N., Subjective Bayesian Methods for Rule Based Inference Systems, National Computer Conference Proceedings, 1976.

[7] Levitt, T.S., Kirby, R.L., and Muller, H.E., A Model Based System for Force Structure Analysis, Proceedings of the SPIE Applications of Artificial Intelligence II, 1985, pp. 169-175.

[8] Mamdani, E.H., and Gaines, B.R., Fuzzy Reasoning and its Applications, New York, 1980.

[9] Marquina, N., Rule-based Evidence Accrual System for Image Understanding, Proceedings of the 1984 SPIE International Conference on Intelligent Robots and Computer Vision, Cambridge, MA, November 1984, pp. 170-175.

[10] McDermott, D., and Doyle, J., Non-Monotonic Logic I, Artificial Intelligence, Vol. 13, April 1980.

[11] Nazif, A., and Levine, M., Low Level Image Segmentation: An Expert System, IEEE Pattern Analysis and Machine Intelligence, September 1984, pp. 555-557.

[12] Negoita, C.V., Expert Systems and Fuzzy Systems, Benjamin/Cummings Publishing, Menlo Park, 1985.

[13] Prade, H., A Computational Approach to Approximate and Plausible Reasoning with Applications to Expert Systems, IEEE Pattern Analysis and Machine Intelligence, May 1985, pp. 260-283.

[14] Rosenfeld, A., The Fuzzy Geometry of Image Subsets, Center for Automation Research, University of Maryland, College Park, July 1983.

[15] Shortcliffe, E.H., and Buchanan, B.G., A Model of Inexact Reasoning in Medicine, Mathematical Biosciences 23, 1975.

[16] Whillock, R., Bazakos, M., and Marquina, N., A Knowledge Based Image Segmentation System, Paper presented at the National Aerosace and Electronics Conference, Dayton, OH, May 1985.

[17] Yager, R., Approximate Reasoning as a Basis for Rule-Based Expert Systems, IEEE Systems, Man, and Cybernetics, July/August 1984, pp. 636-642.

[18] ___, Fuzzy Subsets with Uncertain Membership Grades, IEEE Systems, Man and Cybernetics, March/April 1984, pp. 271-275.

[19] Zadeh, L., Theory of Fuzzy Sets, Encyclopedia of Computer Science and Technology, Marcel Dekker, New York, 1977.

[20] ___, Approximate Reasoning Based on Fuzzy Logic, Proceedings of the International Joint Conference on Artificial Intelligence, 1979, pp. 1004-1010.

[21] ___, Commonsense Knowledge Representation Based on Fuzzy Logic, IEEE Computer, October 1983, pp. 61-65. (special issue on knowledge representation)

[22] ___, Making Computers Think Like People, IEEE Spectrum, August 1984, pp. 26-32.

END

FILMED

11-85

DTIC